Authentication in Health Services

Doctoral Dissertation by

Kirsi Helkala

Submitted to the Faculty of Mathematics and Natural Sciences at the
University of Oslo in partial fulfillment of the requirements for the degree
Philosophiae Doctor (PhD) in Computer Science

AUGUST 2010

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 986

ISSN 1501-7710

Cover: Inger Sandved Anfinsen.
Printed in Norway: AiT e-dit AS.

Produced in co-operation with Unipub.
The thesis is produced by Unipub merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

To Nils

Abstract

The health sector has, as many other sectors in our society, undergone an electronic revolution over the last two decades. However, it differs from others by the sensitivity of medical information, the complexity of the system, and the large number of users. Before the electronic aids each general practitioner and hospital had their own paper patient records. Now these paper records have been transformed to electronic versions and merged together in large databases. Historically, legal restrictions in Norway stated that registers could not be shared electronically. Therefore there were several collections of electronic medical records kept by each practice and each (united) hospital. During the fall of 2009 this law changed, enabling the creation of a national electronic medical record register. Such a register would be available for all medical workers nationwide. The register would provide quick access to patient records in a time of emergency, potentially saving lives. However, simultaneously the risk that sensitive medical information is exposed gets higher. Therefore such a register needs to be strongly protected. Access control and authentication play an important part of this information protection.

This thesis contributes by offering new knowledge on the topic of authentication in health services. The overall goal of the research has been to learn how authentication is done in health services, to point out possible places for improvements, and to develop new authentication mechanisms or enhance existing ones in such a way that they become more secure and user friendly. However, replacing an old authentication mechanism with a new one is not an easy task. A key question in this context is how we know that the new method really is better than the old one?

Different properties of different authentication alternatives make a selection of authentication products difficult. Often, selection of a product is done among similar types of authentication products and is based only on cost or security of the product. However, there are other issues that should be taken into account. For example, strict hygiene standards in operation rooms do not allow smart card use since these are not sterile. On the other hand, passwords can be used because users wear gloves and all equipment including terminals in operation rooms is sterilized. This is a good example showing why the usage scenario should be one of the main selection criteria. This thesis presents a method for authentication product ranking. The method can be used to rank an extensive variety of authentication products (passwords, biometrics, and tokens), and is novel in its provision of a straight forward strategy and of formulas for product ranking. In order to develop the ranking method, the knowledge of authentication methods in a wide range of applications had to be gathered. Therefore, three additional research questions were approached: To what extent are the assumptions on the use of and need for authentication in the health sector valid, is it possible to rank and strengthen methodologies for user generated passwords, and is it possible to identify biometric methods particularly suitable for the health service environment?

The surveys confirmed our hypothesis regarding authentication methods used in health services. Traditional passwords were the most common authentication mechanism and biometrics was not in use at all in 2005. In addition, health enterprises similarly to other enterprises have little knowledge of alternative authentication methods. Our findings confirm that there is a need for ranking methods.

During the thesis work passwords have kept their popularity among authentication products. However, password authentication suffers from the poor quality of human generated passwords. This is not always the user's own fault because guidance for generation of good passwords is seldom given to users. Health enterprises similarly to other enterprises do provide criteria for acceptable passwords. However, these general criteria are too broad to be useful for all users and even if a password fulfils the password criteria, it can still be weak. This thesis addresses these two weaknesses by providing new guidelines for three different password categories separately. Passwords can be divided into Word, Mixture and Non-word password categories. Among these categories users are able to use any mnemonic techniques and generate memorable, but yet strong passwords. This thesis also presents a password quality measurement tool which is a useful aid in a password generation process. As demonstrated in this thesis, the tool and password generation guidelines can be used together to teach password security. Further, an analysis of the education effect is documented.

In addition to the above, this thesis investigates two possible methods of biometrics that could be suitable for the health service environment. As already noted, hygiene is an important issue in health services, and biometrics that is suitable for distance authentication is hygienic. This thesis shows a preliminary work of the performance of a new behavioural biometric, gait. This thesis also presents a suggestion for the improvement in face recognition. The results indicate that the use of only one colour layer (RGB) instead of just a grey image could improve the recognition accuracy.

Acknowledgements

This thesis is dedicated to my husband Nils. If it had not been for you, I would have not even considered starting the whole process. You made me move away from all familiarities and jump into a totally unknown environment. By believing and supporting me, you have shown to me that we really can do anything we put our minds into. Accomplishing two PhDs within the 6 years I have lived here in Norway and still being able to have a wonderful family life with two kids proves to me that we are a good team. Thanks to my family in Finland. You have always stood beside me and my decisions no matter how they have affected you. It has not been easy to have a daughter and sister living not as close by as other members of our tight family. You are the best. Thanks to my mother-in-law for all your support. And many thanks to my children Erik Elmeri and Inga Sofia for showing everyday that life is so much more than work.

The research has been carried out at Norwegian Information Security Lab (NISlab), Gjøvik University College (GUC), Norway. The founding for the research has been provided by the Research Council of Norway [1] Thanks to GUC for providing me a good working environment at college and the possibility to work at home.

There are several people at GUC who I would like to thank for making it possible for me to complete my thesis. First of all I would like to thank my supervisor, Professor Einar Snekkenes for valuable feedback and contributions for my work and being supportive and understanding in cases when family comes first. I also thank Professor Chunming Rong at University of Stavanger for being my second supervisor. I am grateful for Associate Professor Erik Hjelmås for creating a good working condition by making it possible for me to work mostly at home. I am also thankful for Associate Professor Patrick Bours for all the help and co-operation we have had during my thesis period. Lots of thanks to researcher, Dr. Davrondzhon Gafurov, and Torkjel Søndrol for interesting discussions and co-operation. Thanks to Professor Stephen Wolthusen for intellectual feedback and material support in form of chocolate that helped me writing in strange hours.

In addition, advices and comments provided by Adjunct Professor Jan Arild Audestad, Professor Slobodan Petrovic, Adjunct Professor Bernhard Markus Hm̈merli, Professor Christopher Bush, Adjunct Professor Jose J. Gonzalez, and Associate Professor Katrin Franke are appreciated.

Many thanks to Dr. Geir Arne Hjelle for proof reading the thesis.

And last but not last I thank Hanno, Lasse, Geir Olav, Janne, Knut and Frode for many discussions both within and on the outside of the research areas. I have enjoyed the time spent with you also outside of the college.

Kirsi Marjaana Helkala
Vestre Gausdal, August 2010

[1] Grant number 158777/530, “Authentication in a health service context”.

List of Papers

- Surveys in the health services
  - Kirsi Helkala. **Authentication in Norwegian Health Services**. In *Proceedings of the International Symposium on Health Informatics and Bioinformatics*, Turkey '07, 2007
  - Davrondzhon Gafurov, Kirsi Helkala, and Nils Kalstad Svendsen. **Security Models for Electronic Medical Records**. *Telektronikk*, 101 (1), 2005
- Use of biometrics
  - Davrondzhon Gafurov, Kirsi Helkala, and Torkjel Søndrol. **Gait Recognition Using Acceleration from MEMS**. In *Proceedings of The First International Conference on Availability, Reliability and Security (ARES 2006)*, pp. 432-439, 2006
  - Davrondzhon Gafurov, Kirsi Helkala, and Torkjel Søndrol. **Biometric Gait Authentication Using Accelerometer Sensor**. *Journal of Computers* Vol.1, Issue 7, 2006
  - Patrick Bours and Kirsi Helkala. **Face Recognition Using Separate Layers of the RGB Image**. In *Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Special Session on Biometrics - From Sensors to Standardization*, August 15-17, 2008, Harbin, China, pp. 1035-1042, IEEE Press, 2008
- Strengthening passwords
  - Kirsi Helkala and Einar Snekkenes. **Password Generation and Search Space Reduction**, *Journal of Computers*, Vol. 4, Issue 7, pp. 663-669, Academy Publisher, Finland, 2009
  - Kirsi Helkala. **An Educational Tool for Password Quality Measurements**. In *Proceedings of Norwegian Information Security Conference (Norsk Informasjonssikkerhetkonferanse) (NISK2008)* , pp. 69-80, Tapir Akademisk Forlag, 2008
  - Kirsi Helkala. **Password Education Based on Guidelines Tailored to Different Password Categories**. Submitted for publication 2010
- Ranking authentication products
  - Kirsi Helkala and Einar Snekkenes. **A Method for Ranking Authentication Products**. In *Proceedings of the International Symposium on Human Aspects of Information Security & Assurance (HAISA 2008)*, July 8-9, Plymount, UK, pp. 80-93, 2008.
  - Kirsi Helkala and Einar Snekkenes. **Formalizing the Ranking of Authentication Products**. *Information Management & Computer Security - Special Issue*, Vol. 17, Issue 1, pp. 30-43, Emerald, 2009

Thesis Structure

This thesis consists of five parts. Part I is an introduction to the personnel authentication in a health services and a summary of the thesis contributions. Chapter 1 introduces the topic and motivation of the thesis. Chapter 2 outlines the thesis by lightening relationship of the papers. Chapter 3 gives the overview for the personnel authentication. Short summaries of the papers are shown in Chapter 4. Chapter 5 lists the main contributions of the thesis. Future work and conclusion are presented in Chapter 6.

All results of the research work have been presented in previewed conference and journal publications. The papers are grouped based on their topics and shown in the remaining Parts II-V of the thesis. Part II contains two surveys in the health services, Part III contains three publications related to biometrics, Part IV contains three password security related publications, and Part IV contains two publications related to the ranking of authentication products.

Contents

Part I Introduction

1 Introduction 3

2 Thesis Outline 5
2.1 Validity of the Assumptions on the Use of and Need for Authentication in the Health Sector 6
2.2 Identifying Biometric Methods Particularly Suitable for the Health Service Environment 7
2.3 Strenghtening User Generated Passwords 8
2.4 A Method for Ranking Authentication Products 9

3 Personnel Authentication 11
3.1 Knowledge Based Methods 11
3.2 Token Based Methods 14
3.3 Biometrics 15
3.3.1 Fingerprint Recognition 18
3.3.2 Face Recognition. 19
3.3.3 Voice Recogntion 19
3.3.4 Gait Recognition. 20

4 Summary of the Papers 21
4.1 Surveys in the Health Services 21
4.1.1 Security Models for Electronic Medical Records 21
4.1.2 Authentication in Norwegian Health Services 22
4.2 Use of Biometrics 22
4.2.1 Gait Recognition Using Acceleration from MEMS 22
4.2.2 Biometric Gait Authentication Using Accelerometer Sensor 22
4.2.3 Face Recognition Using Separate Layers of the RGB Image 23
4.3 Strengthening Passwords 23
4.3.1 Password Generation and Search Space Reduction 23

4.3.2 An Educational Tool for Password Quality Measurements . . 23
4.3.3 Password Education Based on Guidelines Tailored to Different Password Categories 24
4.4 Ranking Authentication Products 24
4.4.1 A Method for Ranking Authentication Products 24
4.4.2 Formalizing the Ranking of Authentication Products 25

5 Summary of Contributions 27
5.1 Observations on Authentication Requirements and Practices in Health Services 27
5.2 Biometrics 27
5.3 Strengthening User Generated Passwords 28
5.4 A Method for Ranking Authentication Products 28

6 Conclusion and Further Work 31
6.1 Future Work 31
6.2 Conclusion 32
Bibliography 33

Part II Surveys in the Health Services

7 Security Models for Electronic Medical Records 39

8 Authentication in Norwegian Health Services 55

Part III Use of Biometrics

9 Gait Recognition Using Acceleration from MEMS 67

10 Biometric Gait Authentication Using Accelerometer Sensor 75

11 Face Recognition Using Separate Layers of the RGB Image 87

Part IV Strengthening Passwords

12 Password Generation and Search Space Reduction 99

13 An Educational Tool for Password Quality Measurements 109

14 Password Education Based on Guidelines Tailored to Different Password Categories 125

Part V Ranking Authentication Products

15 A Method for Ranking Authentication Products 137

16 Formalizing the Ranking of Authentication Products 151

Part I
Introduction

Chapter 1
Introduction

Since the 1990's our society has undergone an electronic revolution, and literally no segment of the society is left untouched by the possibilities given by computer applications and network services. The health sector is no exception. However, the health sector differs from others by the sensitivity of medical information handeled in its information systems. This is emphasised by strict legal restrictions. The system itself is complex due to the wide range of subsystems and the large number of users. The system users come from both private and public sectors, creating a tight connection between these health sectors. The Norwegian legislation [17, 18, 30–35, 40, 43] defines the legal framework of information exchange and storage. Information security policies, which are based on the legislation, general security standards [36], and specific health security standards [6, 76] are made to guarantee confidentiality, integrity, accountability, and availability for these documents. Despite the good intentions, failures do happen. For example, paper versions of patient journals have been exposed to others [49], patients' electronic medical records have contained medical pictures belonging to other patients [23], and data collapse has deleted medical deliveries [57].

Historically, legal restrictions in Norway stated that medical registers could not be shared electronically. Therefore there were several collections of electronic medical records kept by each practice and each (united) hospital. During the fall of 2009 this law changed, enabling the creation of a national electronic medical record register. Such a register would be available for all medical workers nationwide. The register would provide quick access to patient records in a time of emergency, potentially saving lives. However, simultaneously the risk that sensitive medical information could be exposed gets higher. This has raised concerns about misuse of the common register [85]. Therefore such a register needs to be strongly protected. Personnel authentication and access control play important roles when considering rightful users' access to medical information.

Authentication is about verifying a claimed identity. That is, a prover is aiming to convince a verifier that he is who he claims he is. There are several authentication methods, and they can be categorized. Smith [75] classifies authentication methods in three factors: *something you know* such as a password or a PIN code, *something*

you have such as a magnet stripe card or a mechanical key, and *something you are* such as a fingerprint or a face. After successful authentication, the access control mechanism grants the privilege to use the system with the rights given to that specific user.

Despite all other authentication alternatives, passwords are the most often used authentication method when accessing electronic medical records by a health worker in Norway. Similar situations can be found in the EU area for instance in the UK and in the Andalucian region in Spain [83]. Even though passwords are popular, there are problems with their use. One thing is the poor quality of human generated passwords, while other problems are related to secure handling of them, for example writing them down on visible notes [48] or using the same password in different applications [13].

As a preliminary action towards stronger authentication one could suggest to replace passwords with some other, more secure authentication mechanism. However, enterprises have very little knowledge about the authentication alternatives [27]. Also, the selection task is difficult when one does not know what kind of comparison parameters to include. The usage scenario is as important a criteria for selection as any other criteria, for example cost or security. This was noticed in St. Olav Hospital where selected MDA (hospital version of PDA) did not work as it should in a real situation [11].

The overall goal of this research is to learn about authentication in health services, its weaknesses and possibilities, to suggest improvements, and to develop new authentication mechanisms or enhance existing ones in such a way that authentication becomes more secure and user friendly. This thesis contributes by offering new knowledge on the topic of authentication in health services.

Chapter 2
Thesis Outline

It is easy to say that strengthening authentication can be done simply by replacing the old authentication scheme with another, more secure alternative. But in reality, it is not only security which counts. For example, one-time codes could be used in addition to a static password like done in bank account logons. This adds security, but takes longer time. In an emergency situation a health worker does not have time to type several codes when accessing the electronic medical records (EMR). Fingerprint systems would provide good security, but often they do not handle a situation where a user has gloves on. The essential question in this context is

- How to choose the most suitable authentication product?

Our goal became to develop a method which would help system managers to choose the right authentication product. O'Gorman made a preliminary suggestion regarding what to take into consideration when authentication methods are compared [60]. However, his method is general and focus only on authentication methods, while there really is a need for comparison of actual products. The previous examples also show that the product selection should be done based on the real authentication environment and properties of the user group.

In order to develop the product ranking method, knowledge of authentication methods in a wide range had to be gathered. Therefore, three additional research questions were approached

- To what extent are the assumptions on the use of and need for authentication in the health sector valid?
- Is it possible to identify biometric methods particularly suitable for the health service environment?
- Is it possible to rank and strengthen methodologies for user generated passwords?

Figure 2.1 shows the relations between the main research questions. The relations between papers among each approach are shown in Figure 2.2.

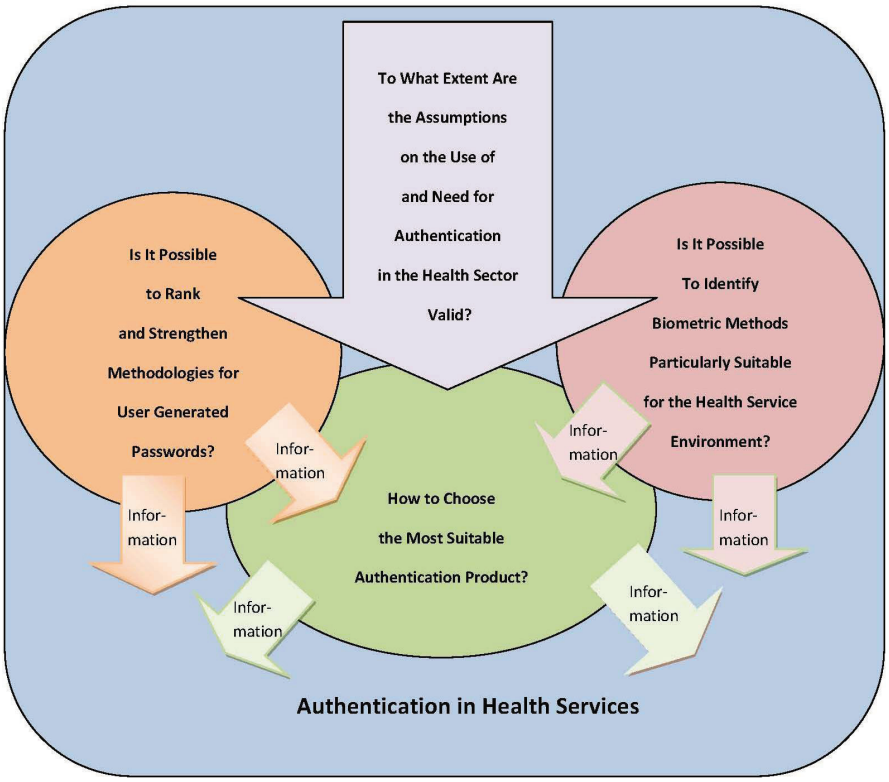


Fig. 2.1 Relations between the main research questions.

2.1 Validity of the Assumptions on the Use of and Need for Authentication in the Health Sector

Surveys were carried out to find answers to the following questions

- Which kinds of authentication mechanisms are in use?
- What are the vulnerabilities of those mechanisms considering also the applications used?
- What kind of framework does the legislation give for the authentication and access control?

Answers to the questions are documented in [22, 25]. The first paper [22] shows the legal aspects of the authentication, models of the information flow and presents alternative solutions to access control in the health services. The second paper [25] mainly concentrates on the relations between health enterprises and authentication methods used in the health services. The surveys also contribute towards the authentication product ranking method by providing concrete knowledge of the usage

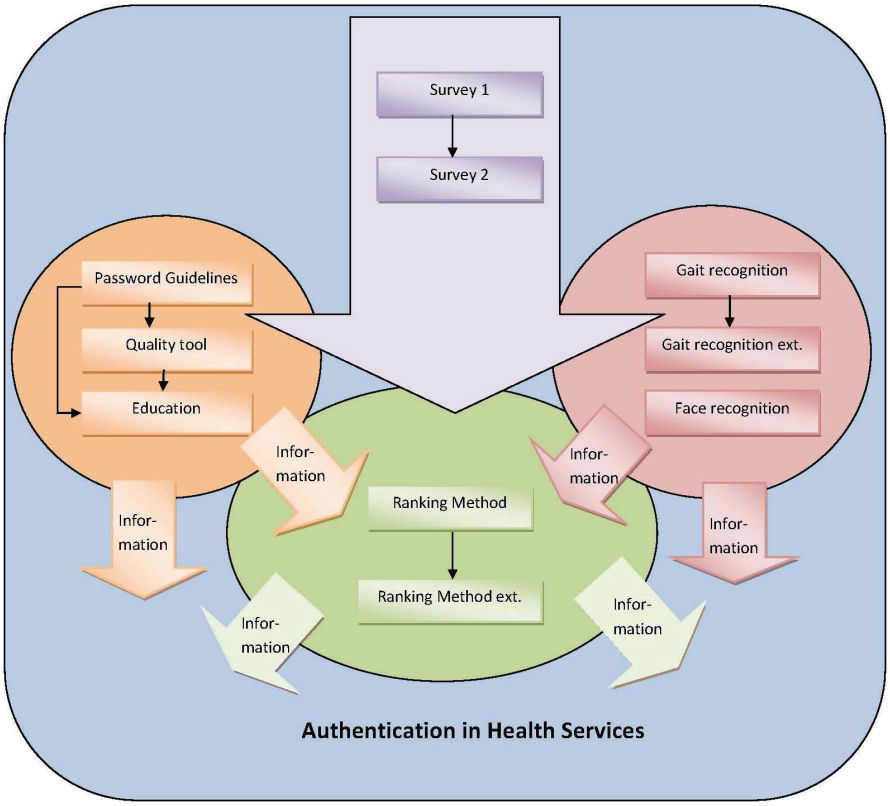


Fig. 2.2 Relations between papers.

scenarios, security mechanisms, security awareness, user behaviour, and needs in the health sector.

2.2 Identifying Biometric Methods Particularly Suitable for the Health Service Environment

The survey findings confirmed that no biometrics was used in the investigated health services. However, products based on biometrics are coming more and more to authentication product markets and they should not be forgotten when considering the alternative authentication products. Therefore, it was decided to investigate two possible methods of biometrics that could be suitable for the health service environment.The methods should be hygenic, implicit for the user, and benefit from health workers dayly rutines. Based on these criterions gait and face recognition were selected.

In [21] we analyse the performance of two gait recognition algortihms (histogram similarity and cycle length) in terms of Equal Error Rate (EER). Paper [20] is an extended version of [21]. The paper [5] presents a face recognition study where we analyse the recognition performance when separate color layers of an RGB image are used for recognition instead of the converted greyscale images. Both performance studies contribute towards the ranking method by providing knowledge of the comparison criteria of biometrics.

2.3 Strenghtening User Generated Passwords

A password scheme was the most common authentication method within the health services in 2005, and during the thesis work it has not lost its popularity. The simple reason for this is that the password system is cheap, easy to implement, maintain, and use. However, the password scheme has weaknesses. Passwords are criticised for providing a poor security. Most of the criticism relates to the user by saying that human generated passwords are weak, easily guessable words which are often stored on the desk beside the terminal. Social engineering and phishing attacks are examples of attack types which are rather easy to carry out in order to steal passwords. These are also the attack types which can be defended against by good information security awareness education. Sasse et. al. [71] points out that the users should be educated to design stronger passwords and they should be given the time and tools needed for the password generation process. However, the proper education is often not provided. To address this, we concentrated on the user perspective of the password security. Our research questions regarding password security were

- How much information will an adversary gain by knowing the password generation guidelines?
- Has education any effect on password strength and memorability?

Paper [29] gives answers to the first question by analyzing the search space reduction of several common password sub-structures within three password categories: Word, Mixture and Non-word passwords. Based on the analysis, new guidelines for a good password generation in each category are introduced. Paper [26] presents a password quality measurement tool which is a follow up of the previous analysis. Both the tool and guidelines are further used to in a password education study which answers the second password research question. The effect of education is reported in [24]. All these studies also provide knowledge of comparison criteria of knowledge based authentication methods for the ranking method.

2.4 A Method for Ranking Authentication Products

For quantitative product ranking, a method should be able to compare and rank authentication products across the authentication categories, include usage scenarios to the selection process, and provide a straight forward strategy and formulas for the product selection. Paper [27] presents such a novel method. Examples of the use of the method are targeted for the security managers in [27]. The usage range of the method is widened to also contain product developers in [28].

Chapter 3
Personnel Authentication

Authentication is about verifying a claimed identity. That is, a prover is aiming to convince a verifier that he is who he claims he is. We consider a setting where the prover is a human, and the verifier may be either a machine or a human. There are several authentication methods, and they can be categorized as follows

- knowledge based methods e.g. a password or a passimage
- token based methods e.g. a smart card or a mechanical key
- biometrics e.g. iris or gait.

Each category is presented in the following sections. Examples of different methods in each category are given, and advantages and disadvantages of the methods are discussed.

3.1 Knowledge Based Methods

The most commonly known knowledge based methods are character string passwords and personal identifying numbers (PIN). A PIN code contains four digits resulting in only 10 000 possible PIN code combinations. This leads to poor security and therefore PIN codes are not used alone, but together with some other authentication mechanism, often a card. Passwords, however, can contain all keyboard characters: letters, digits and special characters. The number of possible characters may be higher than 100 depending on the keyboard used. Password length can be limited by the system and in the most cases minimum length is around eight characters.

Humans have difficulty to remember random character strings. This results in poor habits, for instance dictionary words, password reusing, or passwords to be written down [44, 71]. Different techniques to design memorable yet strong textual passwords have been proposed. An example is a mnemonic password where a long sentence is transformed by different substitution, rotation and removing patterns to a meaningless string of characters [44, 84].

Instead of changing users' password design techniques, a change of authentication scheme has been proposed. In challenge-response protocols a user and machine share secret(s) and the user's identity is verified by the correct answers to the random challenges based on the secret(s). Cognitive and associative password schemes [8] are examples of the challenge-response textual passwords. In a cognitive scheme a user answers several fact and meaning based questions. If a certain number of the answers matches the user's template answers, access to the system is gained. In an associative scheme a user is given words instead of questions. A certain number of correctly associated "answer" words will provide access to the system. The security of these schemes lies on the number of the questions and associative words. The more questions, the more secure the system is. However, a long list of questions takes a long time to answer, leading to poor user friendliness. The challenge-response can be also done by for example numerical computations like in [47] or by visual authentication schemes like in [53].

Visual authentication mechanisms are based on the idea that humans perform better when they are asked to recognize an object or image than when they are retrieving a password string from memory [7]. Examples of these recognition-based systems are random image recognition: Deja Vu [12], graphical password: Draw a Secret [39], and cognitive authentication scheme relying on a secret set of pictures [81].

The traditional password system is cheap to implement. Earlier, most of the password costs came from help desk calls regarding forgotten passwords. Nowadays, the system for resetting passwords has been automated and cost for humans being in-a-loop has been reduced. Passwords are therefore cheap and convenient for both system administrators and users.

Passwords can be stolen from a user with or without the user's co-operation. In a social engineering attack a user is personally contacted by a person claiming to have the right to get the user's password. If emails and/or fake websites are used in order to steal passwords the attack is called a phishing attack. In a phishing attack a user receives an email convincing the user to send private information by email or asking the user to click on the website link in order to authenticate to the application the user tends to use. The opening website is very similar to the original logon site. Untrained people are easily conned with these kinds of websites. An example of this is a successful phishing attack in January 2010 where carbon credits worth of 3 million Euros were stolen [14]. The success rate of these kinds of attacks can be decreased by training people to indentify such threats. The anti-Phishing Working Group's Public Education Initiative [1] is an example of available education. Technical solutions against phishing attacks also exist [2].

Passwords can also be stolen by eavesdropping attacks. A key logger works locally in a computer catching usernames and passwords when they are typed on the keyboard. Local key logger software is possible to detect with antivirus programs or firewalls. Detection of a hardware key logger usually requires physical inspection. Network traffic surveillance software is more difficult to detect. The typed text can also be recovered without any traffic surveillance or key logging software. An attack using keyboard acoustic emanations [87] is an example of this.

Attacks aiming to crack passwords are carried out either on-line or off-line. An off-line dictionary attack requires that an adversary has access to the hashed password file. This is possible in particular circumstances that are unlike to happen in a well-configured systems [7]. An example of such a case is where the computer system has already been thoroughly compromised by an adversary. In this case the adversary is already in position to control the entire system and has no immediate need to break the password file [7]. On the other hand, when considering that users have the tendency to reuse their passwords, the adversary can benefit from cracking the password file and use compromised passwords in an attack targeted to other systems.

Let us assume that the adversary has the hashed password file. An off-line attack continues then by encrypting a special list of guesses with a substitute password mechanism with optimised speed [7]. An encrypted version of each guess is then compared to the hashed password file. When a match is found, the password is recovered. In a basic dictionary attack the list of guesses consists of pure dictionary words. A bit more advanced dictionary attacks have included often used password structures to their own attack dictionaries for instance dictionary words ending a digit. Even more advanced dictionary attacks take advantage of language properties by using the letter combinations which are the most probable in a certain language [55]. These kinds of attacks can also break mnemonic passwords. An attack which goes through every possible combination of characters is called a brute force attack. The defence against these attacks is to force the adversary go through large number of password combinations. This can be done by encouraging users to use characters from all character sets including special characters [75] and by discouraging users from generating passwords with processes that have a low entropy, for example using only words or digits.

In a precomputed dictionary attack a database of all possible passwords along with their encryptions is first created [7]. Then the hashed password file is stolen before a password can be recovered by matching its encrypted version to an entry in the created database. The creation of the database is the most time and memory consuming part of the attack. Earlier it has been too costly in both time and memory for this type of attack to be practical [59]. However, this is not the case anymore since computing power has increased and storage cost has decreased radically. In 2003 Oechslin proposed a new way to precalculate and store the data which significantly reduced the number of calculations during cryptanalysis [59]. As an example he implemented an attack on MS-Windows password hashes and were able to crack 99,9% of all alphanumerical passwords hashes in 13,6 seconds. Currently some precalculated databases together with cracking programs using the tables can be found on Internet [58, 68]. The tables are often called Rainbow Tables after Oechslin.

Rainbow tables are special made to target an attack against certain character sets, password lengths and a hash algorithm. From the user perspective, the only defence is to keep the password search space as large as possible by generating passwords from all character sets including special characters and keeping them long. At the moment, publicly available Rainbow tables [58, 68] do not support the use of special characters in long passwords. From the system perspective, one defence

method against the precomputed attacks is to make it infeasible to calculate all possible password encryptions. This is made by using a salt in encryption. When Oechslin made his example attack, MS-Windows passwords were not encrypted together with salt [7]. Another method type is password stretching which also forces an adversary to compute large tables for comparison. In these methods, a password is strengthened (stretched) by using salt as a second variable together with a password in a strengthening function. User specific salt has been proposed to used instead of known salt [45].

In an on-line attack guesses are submitted to the authentication mechanism. Because attack operates with the pace of the authentication mechanism itself, it is less powerful than off-line attacks [7]. On-line guessing can be defended by an account locking policy. Here an account is locked after certain number of login attempts. However, the use of this mechanism in a network environment might cause problems. For example, large Internet services such as MyYahoo and eBay ended up having problems with denial of service [67].

Computationally harder encryption can protect against both on-line and off-line attacks [7]. In these methods the password encryption in an authentication process is slowed down making an attack time consuming.

3.2 Token Based Methods

In token based methods a user has to have a token in his possession in order to authenticate to the system. Based on the tokens role in the authentication process or protocol, tokens fall into two categories: passive and active tokens [75]. A passive token is a storage device for the secret. Examples of passive tokens are mechanical keys and magnetic stripe ID-cards or ATM cards. An active token contains an electronic chip or micro computer and is therefore able to generate different outputs under different circumstances. Examples of active tokens are one-time password generators and smart cards. A one-time password generator contains a six or eight digit number. The number is generated, for example with a unique token identifier and a time signal. The number changes within a certain time period, for example every one to three minutes. A smart card goes through a challenge-response protocol with a reading terminal before the secret is released. In addition to the previous categories, tokens can be divided into two sets based on what kind of contact they need with a verifier machine. An example of a token which needs direct contact is a mechanical key. Contactless tokens work from the distance, for instance cards used for accessing lifts in the slalom slopes.

Tokens are portable and their absence is usually noticed. Therefore, a loss or theft can be quickly reported. In a case of loss, theft or malfunction, a new token can be issued by the system owner. However, it takes longer time to receive a new token than to reset a new password. Normally a new token can be gotten in a couple of days. Tokens especially cards, are often used together with PIN-codes to achieve a two-factor authentication.

Passive tokens are often rather easy to copy without the user noticing. Examples of these are payment card forgeries when a copy machine is installed either on ATM machines or in payment terminals. The latter was used in a large chain of card forgeries in several countries in Europe 2008 [42]. A copy of a key has classically been made by using mould, but alternative approaches can also be found. For example, keys have been copied from a picture [79].

Active tokens can use cryptographic techniques and therefore be secured against sniffing and replay attacks [75]. Smart cards are more robust against forgeries because their challenge-response protocol does not work in a unauthorized card terminals. However, they are not unbreakable. Probing attacks [73] and fault generation attacks [46] are examples of attacks which need possession of a token. Probing attacks are done on the chips surface while in fault generation attacks abnormal environmental conditions are introduced to produce erroneous behaviour which then leaks information. Differential Power Analysis (DPA) and Simple Power Analysis (SPA) [41] are examples of eavesdropping attacks. Power analysis attacks are non-invasive, and they exploit externally available information. DPA allows extraction of secret keys by analyzing power consumption. Simple Power Analysis is a simpler form of the attack that does not require statistical analysis.

3.3 Biometrics

Biometrics is not a new field of methods in authentication. It has been used for centuries [75] and we use it on a daily basis without noticing when recognizing our family members, friends, co-workers from the way they look, talk, walk etc. However, biometrics has been utilized in computer technologies after we have learned how to extract biometric features from a person. Biometric features can be divided into two categories: physiological and behavioural. They are also refered to as stable and alterable features [60]. Examples of the physiological features are fingerprint, face, hand and iris. Voice, gait and keystroke are examples of the behavioural features.

According to [50] any human physiological and behavioural feature can be used to recognize a person as long as it satisfies the following seven requirements: universality, distinctiveness, permanence, collectability, performance, acceptability, and circumvention. An ideal biometric feature for the authentication purposes is universal, meaning that everybody has it. It is unique for each person. It is permanent for a long time period and it is easy to collect from a person. It performs well when processed. People accept its use without hesitation and it is very difficult to forge. Table 3.1 compares fingerprint, face, voice, and gait biometric identifiers along these requirements.

When a person is authenticated with passwords, the provided string should fully match the earlier registered password. If this is not the case, access is denied. This decision strategy does not apply to biometrics. In biometrics, the enrolment sample and sample provided when authenticated are never exactly the same. Environmen-

Table 3.1 The table contains comparisons of the fingerprint, face, voice, and gait biometric identifiers along different requirements. Abbreviations are as follows: H means high, M medium and L low. The table also shows error rates for some specific recognition modes for each biometric identifiers.

	FINGER	FACE	VOICE	GAIT
Universality	M	H	M	M
Distinctiveness	H	L	L	L
Permanence	H	M	L	L
Collectability	M	H	M	H
Performance	H	L	L	L
Acceptability	M	H	H	H
Circumvention	M	H	H	M
Reference	[37]	[37]	[37]	[37]
Description of recogition mode	Two Fingers	Three Dimentional	Long vs Long Same microp.	Sensor on ankle
EER	0,0008%		0,7%	5%
FAR		0,001%		
FRR		0,01%		
Reference	[80]	[66]	[56]	[19]

tal differences might affect the sample quality, for example lightning conditions, temperature, and humidity. Physiological changes in a human body might affect the feature itself, for instance a person having a cold or has injured himself. In addition, errors or noise in sensors like dirt on the sensor device might affect recognition accuracy. Therefore, the decision in biometrics is done within a certain confidentiality limit. If the provided sample is close enough (threshold limit is set up for each system separately), the sample is accepted and the person gains access rights.

Because the samples differ each time, the errors in authentication also occur. There are several different error types in a biometric system [50, 51]. Not all people can provide all biometric features. For example, a blind person cannot provide iris-code [51] or gait feature is impossible for a person sitting on a wheelchair. This leads to the Failure to Enrol Rate (FTE) measuring the proportion of individuals for whom the system is unable to generate templates. Failure to Acquire Rate (FTA) again measures the proportion of attempts for which the system is unable to capture a biometric feature of sufficient quality. An example of this is a person having a cast around his hand when trying to authenticate with hand recognition or a scarf covering the face in face recognition or a dirty sensor causing problems in fingerprint recognition.

The matching errors are done when two samples are compared and the difference is further compared to the threshold value. This causes two kinds of errors: False Non-Match Rate (FNMR) and False Match Rate (FMR) [50]. False Non-Match Rate measures the proportion of the mistaking two biometric samples from the same source to be from two different sources. False Match Rate again measures the proportion of the mistaking two biometric samples from two different sources to be from same source. Figure 3.1 illustrates the computation of FMR and FNMR. Empirically, FNMR and FMR can be estimated as follows

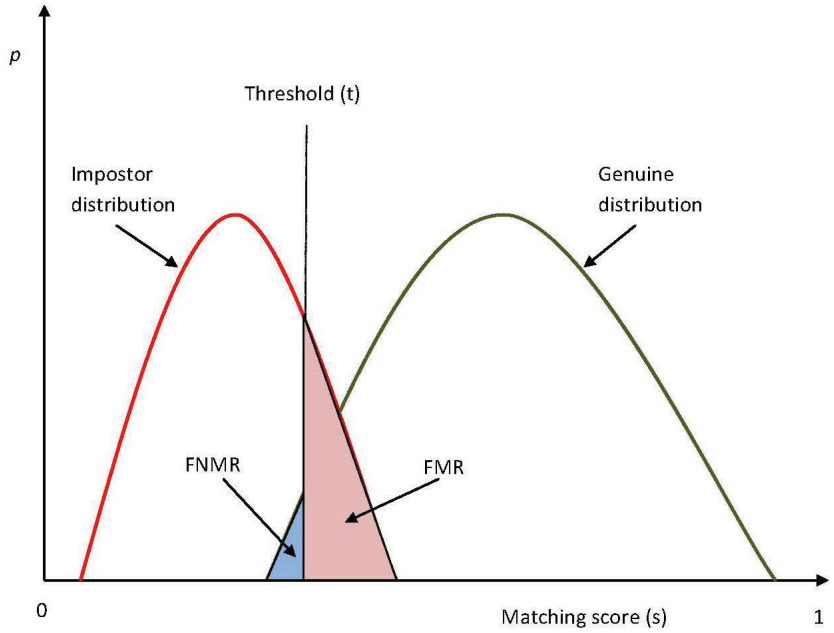


Fig. 3.1 FMR and FNMR for a given threshold t are displayed over genuine and impostor score distributions. Drawn based on [50].

$$FNMR = \frac{Number\ of\ non-matched\ genuine\ attempts}{Total\ number\ of\ genuine\ attempts} \tag{3.1}$$

$$FMR = \frac{Number\ of\ matched\ impostor\ attempts}{Total\ number\ of\ impostor\ attempts} \tag{3.2}$$

The biometric system performance at all threshold points t can be reported by plotting a Receiver Operating Characteristic (ROC) curve [50]. Figure 3.2 shows an example of a ROC curve. The rate when FNMR = FMR is called Equal Error Rate (EER).

When matching errors measure accuracy of the matching process, the decision accuracy is measured by decision errors False Rejection Rate (FRR), and False Acceptance Rate (FAR). These measures combine FMR, FNMR and FTA in accordance with the system policy [51]. Formulas for FAR and FRR in the case when verification decision is done based on a single attempt are as follows:

$$FAR(t) = (1 - FTA)FMR(t) \tag{3.3}$$

$$FRR(t) = (1 - FTA)FNMR(t) + FTA, \tag{3.4}$$

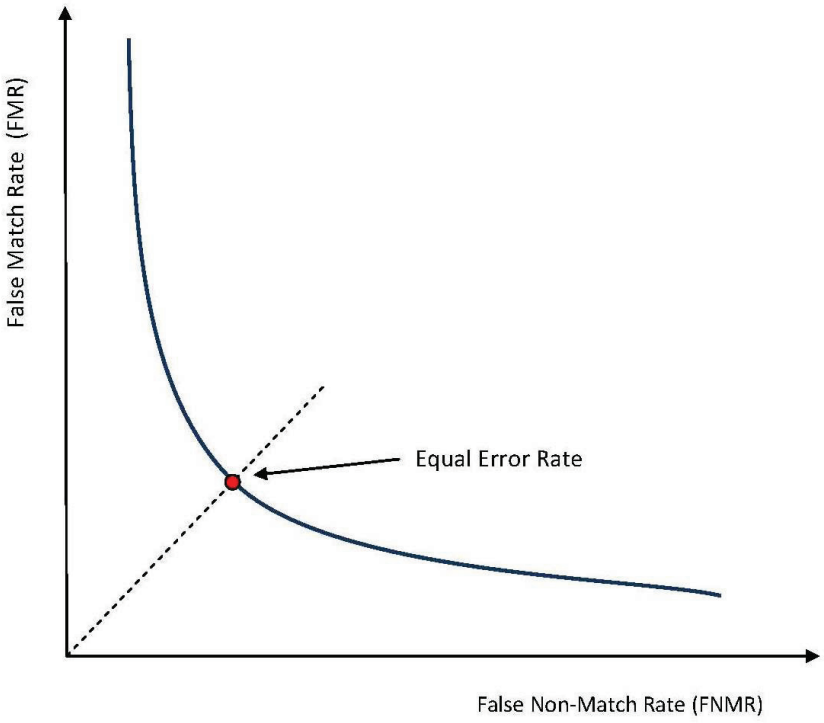


Fig. 3.2 FMR and FNMR tradeoff. Drawn based on [50].

where t is a threshold value. Decision Error Trade-off (DET) curve for FAR and FRR can be drawn similarly to the ROC curve earlier. In cases where only matching errors are taken into account in a verification mode, decision errors become matching errors: $FAR = FMR$ and $FRR = FNMR$. As an example of error rates Table 3.1 shows error rates for some specific recognition modes for fingerprint, face, voice, and gait biometrics.

In general, biometrics are easy to use for authentication purposes because they are always with the person. However, they also have disadvantages. The recognition equipment might be expensive, characteristics can be injured and they cannot be changed. Further, replay attacks are possible and false rejection of legal users can be a problem when security level is high. And last, but not least, the person's privacy might be at risk.

3.3.1 Fingerprint Recognition

Fingerprint recognition was put into use in 1893, when the Home Ministry Office in United Kingdom accepted that no two individuals have the same fingerprints [50].

Even twins have different patterns in their fingertips. The pattern does not change after teenage years. However, there are some ethnic groups which have very invisible fingerprints. Fingerprints can also be torn away, especially among people who work hard with their hands. There are several different sensor types to capture fingerprint image. Capturing can be done by using optical frustrated total internal reflection (FTIR), ultrasound reflection, piezoelectric effect, temperature differential or capacitance differential [50]. Injuries and cuts might alter a fingerprint temporary. Skin elasticity, pressure, oily or dry fingers can effect image quality [50, 69]. Fingerprints can be copied easily and access with an artificial print can be gained [70]. However, liveness detection increases fingerprint recognition security [70].

3.3.2 Face Recognition.

The face is one of the most natural biometric features for recognition purposes. We use it on a daily basis, either recognizing friends or matching ID-documents to the face of the document provider. Image capturing itself is an easy and non-intrusive process. Face recognition has been studied for over 35 years. During this time several face recognition algorithms have been proposed and evaluated [4, 52, 61–65]. Currently face recognition is widely used and according to [86], the three most common methods are: The Elastic Bunch Graph Matching (EBGM) system, the subspace LDA (Linear Discriminant Analysis) system, and the probabilistic eigenface system. Even though face recognition is widely studied, there still remain several challenges. Illumination, poses and facial expressions cause the most severe problems to receive accurate face recognition [86].

3.3.3 Voice Recogntion

Voice is a behavioural biometric because the manner, motion and pronunciation of the words varies [82]. There are two modes for the voice recognition: text-dependent and text-independent mode. In a text-dependent mode, a user says predetermined words. In a text-independent mode the input is free. According to [82] text-dependent is a more accurate recognition mode. The challenges for the voice recognition are changes of the voice pattern over the years. Sickness, emotions and room acoustics can also alter the voice signal temporary [82]. Attack types for the voice recognition systems are mimicry by humans and replay attacks [50]. In a replay attack a recorded voice sample is played. Depending on the quality of the recording device, the provided sample might be hard to separate from the original sample. The voice recognition technology is an interesting authentication alternative for systems where speech is natural behaviour for a user. Examples for such cases are the use of a mobile phones or a case where user's oral notes are transformed to text by a software.

3.3.4 Gait Recognition.

The gait is a person's manner of walking and according to the early medical studies the gait pattern has a high degree of uniqueness [54]. There are three approaches to capture gait pattern. Capturing can be done based on machine vision, floor sensors or wearable sensor systems [19]. A wearable sensor can be installed in a mobile device and therefore it can be thought to be used for device protection. In our studies, we have used wearable sensors. The gait signal varies depending on the sensor placement on a body. With current signal comparison algorithms, the gait signal should be produced around the same body part. Other challenging points are, for example different walking surfaces, different shoes, and difference of carrying and not carrying objects [19]. Being a behavioural biometric, mimicry attacks might be possible [77].

Chapter 4
Summary of the Papers

This chapter contains a summary of the papers in the form of paper abstracts.

4.1 Surveys in the Health Services

Surveys were done to find answers to the research question: *To what extent are the assumptions on the use of and need for authentication in the health sector valid?* The following two papers are the documentation of the survey results.

4.1.1 Security Models for Electronic Medical Records [22]

In this article we give a description of the electronic medical record (EMR) and sumarize the main legal matters related to it. Further, we propose an information flow model of the current Norwegian EMR. This model is used to show that current access rights to the EMR are too general, and that requirements on accountability are not fulfilled. Finally we show how the use of dynamic access rights and non-repudiation can be used to improve the current situation. Role based access control is proposed to limit the access to medical records in internal information exchange, and non-repudiation to achieve accountability for exteranal information exchange. Both solutions require the implementation of a public key infrastructure in the health sector. The proposed solution for external communication is an alternative to the propositions of a centralized health register. Our solution leads to effective communication among institutions, while avoiding the vulnerabilities introduced in collecting all the information in one database.

4.1.2 Authentication in Norwegian Health Services [25]

This paper is a survey report on current authentication methods used in Norwegian health enterprises, both in external and internal information exchange. The paper is based on interviews of health workers working in different health enterprises, as well as IT-workers and programmers in Eastern Health Region in Norway. The interviews took place in November 2004 and March 2005. The surveys showed that the most commonly used authentication method in the internal information maintaining and transfer is a password. The two-factor key cards are used to gain a physical access to the buildings and rooms. No biometrics are in use.

4.2 Use of Biometrics

The studies with biometrics were done to answer to the question: *Is it possible to identify biometric methods particularly suitable for the health service environment?*

4.2.1 Gait Recognition Using Acceleration from MEMS [21]

This paper presents an approach on recognising individuals based on 3D acceleration data from walking, which are collected using MEMS. Unlike most other gait recognition methods, which are based on video source, our approach uses walking acceleration in three directions: vertical, backward-forward and sideways. Using gait samples from 21 individuals and applying two methods, histogram similarity and cycle length, the equal error rates of 5% and 9% are achieved, respectively.

4.2.2 Biometric Gait Authentication Using Accelerometer Sensor [20]

This paper presents a biometric user authentication based on a person's gait. Unlike most previous gait recognition approaches, which are based on machine vision techniques, in our approach gait patterns are extracted from a physical device attached to the lower leg. From the output of the device accelerations in three directions: vertical, forward-backward, and sideways motion of the lower leg are obtained. A combination of these accelerations is used for authentication. Applying two different methods, histogram similarity and cycle length, equal error rates (EER) of 5% and 9% were achieved, respectively. This paper is an expanded version of [21].

4.2.3 Face Recognition Using Separate Layers of the RGB Image [5]

In many cases face recognition of still images is performed with greyscale images. These images are actually converted from a color image to greyscale before the analysis takes place. A consequence of such a conversion is obviously loss of information, which could influence the performance of the face recognition system. It would be interesting to see if using one of the three color layers of the RGB image could give better recognition performance compared to the greyscale converted image. We conducted two experiments and the results indeed support this idea. We found that the red layer of the RGB image gives the best recognition performance, especially in the cases where an extra light source is used to light up (part of) the face of the participants in the experiments. In the case that the participants were facing the camera we saw the Equal Error Rate drop from 3.3% for the greyscale images to 1.8% for the red layer of the RGB images in our initial experiment.

4.3 Strengthening Passwords

The following three papers are focused on finding answers to the research question: *Is it possible to rank and strengthen methodologies for user generated passwords?*

4.3.1 Password Generation and Search Space Reduction [29]

It is easy for humans to design passwords that are easily remembered. However, such passwords may have a predictable structure, making exhaustive search feasible. We have divided human-generated passwords into three categories: Non-word passwords, Mixture passwords, and Word passwords; depending on their overall structure. Within these categories, we have analyzed the search space reduction of several common password sub-structures. From this analysis, we have derived guidelines that yield strong passwords within in each password category. Our results contribute towards the goal of achieving both strong *and* memorable passwords.

4.3.2 An Educational Tool for Password Quality Measurements [26]

Strong and memorable passwords are highly desirable. Password policies are defined to instruct users to design strong passwords, but the education of different password designing methods is often forgotten. And furthermore, the general in-

structions do not apply to all designing processes, and the users' own desires are then ignored. This might result in poor passwords even when the recommendations are fulfilled. As a response to these drawbacks, we created a tool which evaluates the quality of a password while guiding the user through the design of a password. Passwords are categorized in three categories based on their structure information. The structure information is then used to compute a numerical quality score. The tool differs from the password checkers found on the Internet, because it is able to compute the quality score without receiving the actual password string.

4.3.3 Password Education Based on Guidelines Tailored to Different Password Categories [24]

General password policies do not guarantee that passwords fulfilling the requirement are good enough. The policies have a tendency to be too broad to be useful for all users. Different users have different design processes based on what kind of passwords they most easily remember. Users are also often left to generate passwords on their own without any training. In our study we measured the effect of education on the strength of a password. In order to help users to create good passwords, we divided passwords into three password categories: Word password, Mixture password and Non-word password. For each category different password generation guidelines were taught. Participants had access to the password quality measurement tool, which not only measured the strength of the password but also guided students in the generation process. It was shown that education had a positive effect and that passwords became stronger right after the education. The most important result was that a password structure got changed as the variation of structures increased and different structure types were more evenly distributed. However, after half a year without reminders or education repetition, most of the positive effect was lost. While password structures still differed, they had become less complex, in particular as participants had given up using special characters.

4.4 Ranking Authentication Products

The last two papers answer the first research question: *How to choose the most suitable authentication product?*

4.4.1 A Method for Ranking Authentication Products [27]

There is a steady increase in both authentication methods, and in products implementing these methods. Product selection has impact on strategic factors such as

system security, cost and usability. This paper presents a new method for ranking authentication products. The method can contribute towards an improved decision process. Using our method, issues such as technical performance, application/system specific requirements, cost and usability are addressed.

4.4.2 Formalizing the Ranking of Authentication Products [28]

This paper presents a novel method for ranking authentication products. Using our method, issues such as technical performance, application/system-specific requirements, cost and usability are addressed. The method simplifies and makes the selection process more transparent by identifying issues that are important when selecting products. We have used quantitative cost and performance analysis. Our method can be widely applied, allowing the comparison and ranking of an extensive variety of authentication products (passwords, biometrics, tokens). The method can be used for both product selection and the process of product development as supported by the case studies. This paper is an expanded version of [27].

Chapter 5
Summary of Contributions

Here we present the main contributions within each research topic.

5.1 Observations on Authentication Requirements and Practices in Health Services

The surveys pointed out the needs for improvements regarding authentication in health services. The health services would benefit from security education, including topics such as the generation of good passwords, password management, and defence against social engineering attacks. Also, the use of other authentication methods instead of passwords was found to be an interesting idea, especially among the end-users. Being able to authenticate indirectly, meaning that authentication could be handled while carrying out other tasks, is tempting because it would let a health worker concentrate only on the patient's health.

In addition to the need for personnel authentication improvements, we modelled internal information flow within the Norwegian health services by using the Bell-La Padula Confidentiality Model. This helped us point out a violation against legal restriction stating that access to a certain patient record should only be allowed to health workers treating the patient. With an example of role-based access control in EMR, we showed how to correct the violation. Similarly, we showed a thorough example of the use of non-repudiation in external information exchange.

5.2 Biometrics

The gait study showed that it is possible to use gait as an authentication method when recognition is done based on the gait acceleration data. In our studies, the performance of two recognition algorithms of our own (histogam similarity and cycle length) were analysed. The acceleration data was collected by a sensor with a low

sampling rate. A succesful use of low sampling rates implies that the authentication system manages with low resources. Therefore, mobile device protection by gait could be one future application.

Most of the papers on face recognition with eigenfaces use greyscale pictures. In such cases, a camera captures colour images that are transformed into grey images. We showed, however, that it might be better to not transform the images to greyscale, but use the separate colour layers of the original image. In our study, the red layer of the RGB colour image performed best. However, the reason why the red layer performs best is not entirely clear. Explanatory factors can be the light used in our experiments having a large peak in the red area of the spectral graph, skin colour of our volunteers, and the characteristics of the camera.

5.3 Strengthening User Generated Passwords

A secure password is both memorable and strong. What is considered a memorable password differs from person to person. Someone remembers word-related passwords better, others are capable of remembering long meaningless character strings without any mnemonic tactic. In order to be useful for all users, we divided passwords into three categories: Word, Mixture and Non-word passwords and generated password design guidelines for each category.

Further, we developed a password quality measurement tool, which uses same password categories when computing quality score. The tool computes a quality score based on the search space entropy of a certain password structure. The structure information is given by providing numeric answers to the tool's questionnaire. Therefore, the actual password is never revealed like in other tools computing the strength of a password.

Education has a postive effect on users which is very strong right after the education. However, the effect of the education vanishes in the long run if the users do not get reminders of the passwords guidelines. This indicates that continueed tutoring and evaluation should be implemented in the systems using password authentication.

5.4 A Method for Ranking Authentication Products

We developed a method that can help enterprises identify issues that should be considered when choosing authentication products. The method is divided into four main categories: suitability for the user and environment compatibility, security level compatibility, usability and cost. The method eliminates unusable products and lists the remaining products according to their cost. The key point in our method is to turn authentication method property vectors to scalors by using usage scenario requirements as weight factors.

Further, we derived formulas for security level comparison. In our approach, the security level of an authentication method is determined to be the minimum of two entropy measures. The first one is the authenticator's search space, and the second is the difficulty for an attacker to engineer a circumvention attack. The hardness of the circumvention is computed from estimates of the circumvention probabilities without user awareness. Since circumvention strategies differ for the different authentication factors, we defined a separate formula for each authentication category.

Chapter 6
Conclusion and Further Work

6.1 Future Work

The need for strong personnel authentication is growing steadily as new applications having remote access to electronically storaged information are coming to markets. Some of the applications are specially targeted to the health services. Portable devices can be handy in use, but they might also compromise confidential information in the case of theft. Therefore, it would be desired to have an authentication mechanism which would shut down the connection when an unauthorized user is holding the device. Health services also have several special environments where authentication is needed to be done such as operation rooms. Biometrics can provide suitable authentication alternatives for these cases.

- Our gait study was motivated by the idea of mobile phone protection. The idea can be widened to apply other mobile devices, such done in [38, 78]. In the future it would be interesting to see how gait would perform with mobile ICT in hospital setting such as in [10].
- Operation rooms need an authentication mechanism which is hygienic. Face recognition could provide such a benefit. However, chirurgical staff wears masks leaving only the eyes unveiled. Such a partially covered face is a problem in face recognition. In our study we used a strong spot light to illuminate faces, and we found that the best accuracy performance was achieved with the red colour layer. The strong lighting is also in use in operation rooms. It would be worth studying if the use of separate colour layers provides better recognition accuracy in the case of partially covered faces.
- Hospitals and general practices have a possibility to use speech recognition technology which transforms oral text to written text [74]. Voice would be a natural authentication method for such a system.

Previous ideas were related to authentication in a special environment or of an application in the health services. In addition, we would like to outline research topics regarding authentication methods themselves.

- The memorability of passwords is a problem which is hard to overcome. Would it help if there were a password hash algorithm that could give a strong password for each separate site or application? Could it be designed an algorithm usable without any tool which would provide a high quality password from easily remembered input?
- Colour images used in face recogntion left much room for future work. For example:
  - Does the red layer of an RGB image contain the best information because it resembles best the color of the skin for Caucasian people?
  - What combination of color layers performs best in case of non-Caucasian people?
  - The best results for the red layer of the image might indicate that using Near Infrared (NIR) lighting and image capturing could also give good results for face recognition.

6.2 Conclusion

This thesis contributes by offering new knowledge on the topic of authentication in health services. The overall goal of the research has been to learn how authentication is done in health services, to point out possible places for improvements, and to develop new authentication mechanisms or enhance existing ones in a way such that they become more secure and user friendly. One important issue when protecting information systems is to use an authentication product which is the most suitable for the given usage scenario. However, enterprises have little knowledge about comparison of alternative authentication products. To address this, we invented a method for authentication product ranking of an extensive variety of authentication products (passwords, biometrics, and tokens). To our knowledge, the method was the first one which provided strategy and formulas for authentication product ranking. Our ranking method was referenced to in [72], where a new access control method based on the results of the risk assessment was proposed.

In order to develop the ranking method, the research was divided into separate topics. Each of them contributed also individual results. These results have been helpful to others in their research. Survey paper [22] contained both survey results and propositions for improvements. The models of the current systems were referenced to in [15]. The proposed alternative methods were solutions for the situations where the currently used methods did not meet the legal restrictions. Role based access control was proposed to limit the access to medical records in internal information exchange, and non-repudiation to achieve accountability for external information exchange. The role based access control was further referenced to in [16].

Gafurov has continued the gait study in his thesis [19] and together with several co-authors developed the first gait recognition algorithms introduced in [21] further on. In addition, [20] has been referenced to for instance in [3, 9, 38, 78]. For

example, Tamviruzzaman et. al [78] used gait and GSP-location signals to authenticate the owner of an iPhone. In the face recognition study the recognition was done based on the different color layers of the RGB image. Recognition accuracy was improved, but the study left room for future work with the use of color layers in face recognition.

Passwords are still the most common authentication method. We have provided methods for improving their security. We have generated new, more specific guidelines for good password generation. Based on the guidelines we have built a quality measurement tool to help in the password generation process. The tool and guidelines are to be used in security education improving situations like [13, 48] mentioned earlier in the Introduction chapter. However, one-time education is not enough to keep the security level acceptable. In order to achieve good security with passwords, users should get reminders of the good password habits regularly.

Bibliography

1. Anti-Phishing Working Group: APWG Public Education Initiative. `education.apwg.org`. Last visited: 22.3.2010
2. Anti-Phishing Working Group: Sponsor Solutions. `www.antiphishing.org/solutions.html`. Last visited: 22.3.2010
3. Bächlin, M., Schumm, J., Roggen, D., Töster, G.: Quantifying Gait Similarity: User Authentication and Real-World Challenge. In: Proceedings of the Third International Conference on Advances in Biometrics, *LNCS*, vol. 5558, pp. 1040–1049 (2009)
4. Blackburn, D., Bone, M., Phillips, P.: Facial Recognition Vendor Test 2000 – Evaluation Report. Tech. rep., NIST (2001)
5. Bours, P., Helkala, K.: Face Recognition Using Separate Layers of the RGB Image. In: Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Special Session on Biometrics - From Sensors to Standardization, pp. 1035–1042. IEEE Press (2008)
6. British Standard: Health Informatics - Secure User Identification for Health Care - Management and Security of Authentication by Passwords (2004)
7. Brostoff, A.: Improving Password System Effectiveness. Ph.D. thesis, Department of Computer Science, University College London (2004)
8. Bunnell, J., Podd, J., Henderson, R., Napier, R., Kennedy-Moffat, J.: Cognitive, Associative and Conventional Passwords: Recall and Guessing Rates. Computers & Security **16**(7), 629–641 (1997)
9. Czeskis, A., Koscher, K., Smith, J.R., Kohno, T.: RFIDs and Secret Handshakes: Defending Against Ghost-and-Leech Attacks and Unauthorized Reads with Context-Aware Communications. In: Proceedings of the 15th ACM conference on Computer and Communications Security, pp. 479–490 (2008)
10. Dahl, Y., Alsos, O.A., Svans, D.: Evaluating Mobile Usability: The Role of Fidelity in Full-Scale Laboratory Simulations with Mobile ICT for Hospitals. In: Proceedings of the 13th International Conference on Human-Computer Interaction. Part I: New Trends, *LNCS*, vol. 5610, pp. 232–241 (2009)
11. Dahle, D.Y.: IT-skandale ved St. Olavs hospital. `www.tu.no/data/article94383.ece` (2007)
12. Dhamija, R., Perrig, A.: Dej'a Vu: A User Study Using Images for Authentication. In: Proceedings of the 9th USENIX Security Symposium (2000)

13. Ernes, A.K.B.: Bruker samme passord overalt. www.digi.no/807513/bruker-samme-passord-overalt (2009)
14. Espiner, T.: Phishing scam spurs EC into security revamp. news.zdnet.co.uk/security/0,1000000189,40025391,00.htm (2010)
15. Feltus, C.: Preliminary Literature Review of Policy Engineering Methods. In: Proceedings of the third International Conference on Information and Communication Technologies: From Theory to Applications, ICTTA, pp. 1–6 (2008)
16. Feltus, C., Petit, M.: Building a Responsibility Model including Accountability, Capability and Commitment. In: Proceedings of International Conference on Availability, Reliability and Security, ARES, pp. 412–419 (2009)
17. Forsvarsdepartementet: Lov om forebyggende sikkerhetstjeneste. www.lovdata.no (1998)
18. Forsvarsdepartementet: Forskrift om informasjonssikkerhet. www.lovdata.no (2001)
19. Gafurov, D.: Performance and Security Analysis of Gait-based User Authentication. Ph.D. thesis, Faculty of Mathematics and Natural Science, University of Oslo (2008)
20. Gafurov, D., Helkala, K., Søndrol, T.: Biometric gait authentication using accelerometer sensor. Journal of Computers **1**(7), 51–59 (2006)
21. Gafurov, D., Helkala, K., Søndrol, T.: Gait Recognition Using Acceleration from MEMS. In: Proceedings of the First IEEE International Conference on Availability, Reliability and Security (ARES), pp. 432 – 439. Vienna, Austria (2006)
22. Gafurov, D., Helkala, K., Svendsen, N.K.: Security models for electronic medical records. Telektronikk **101**(1), 98–110 (2005)
23. Haaland, L.: Røntgenbilder byttet plass. www.tu.no/nettarkiv/article32517.ece (2005)
24. Helkala, K.: Password Education Based on Guidelines Tailored to Different Password Categories. Submitted to publication
25. Helkala, K.: Authentication in Norwegian health services (survey report). In: Proceedings of International Symposium on Health Informatics and Bioinformatics, Turkey '07. Antalya, Turkey (2007)
26. Helkala, K.: An educational tool for password quality measurements. In: Proceedings of Norwegian Information Security Conference, pp. 69–80. Tapir Akademisk Forlag (2008)
27. Helkala, K., Snekkenes, E.: A method for ranking authentication products. In: Proceedings of the International Symposium on Human Aspects of Information Security & Assurance (HAISA 2008), pp. 80–03 (2008)
28. Helkala, K., Snekkenes, E.: Formalizing the ranking of authentication products. Information Management & Computer Security Special Issue, Emerald **17**(1), 30–43 (2009)
29. Helkala, K., Snekkenes, E.: Password generation and search space reduction. Journal of Computers **4, Issue 7**, 663–669 (2009)
30. Helse- og omsorgsdepartementet: Lov om helsetjenesten i kommunene. www.lovdata.no (1982)
31. Helse- og omsorgsdepartementet: Lov om helsepersonell m.v. (helsepersonelloven). www.lovdata.no (1999)
32. Helse- og omsorgsdepartementet: Lov om pasientrettigheter (pasientrettighetsloven). www.lovdata.no (1999)
33. Helse- og omsorgsdepartementet: Lov om spesialisthelsetjenesten m.m. www.lovdata.no (1999)
34. Helse- og omsorgsdepartementet: Forskrift om pasientjournal. www.lovdata.no (2000)
35. Helse- og omsorgsdepartementet: Lov om helseregistre og behandling av helseopplysninger. www.lovdata.no (2001)
36. ISO: NS-ISO/IEC 27002:2005 (2005)
37. Jain, A.K., Ross, A., Prabhakar, S.: An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics **14**, 4–20 (2004)
38. Jakobsson, M., Shi, E., Golle, P., Chow, R.: Implicit authentication for mobile devices. In: Proceedings of the 4th USENIX Workshop on Hot Topics in Security (HotSec '09) (2009)

39. Jermyn, I., Mayer, A., Monrose, F., Reiter, M.K., Rubin, A.D.: The design and analysis of graphical passwords. In: Proceedings of the 8th conference on USENIX Security Symposium, vol. 8, pp. 1–1 (1999)
40. Justis- og politidepartementet: Lov om behandling av personopplysninger (personopplysningsloven). www.lovdata.no (2000)
41. Kocher, P., Jaffe, J., Jun, B.: Differential Power Analysis. www.cryptography.com/resources/whitepapers/DPA.pdf
42. Kotilainen, S.: Uskomaton jättirikos: kaupoissa rikollisten kortinlukijoita. www.tietokone.fi/uutiset/2008/uskomaton_jattirikos_kaupoissa_rikollisten_kortinlukijoita (2008)
43. Kulturdepartementet: Lov om arkiv (arkivloven). www.lovdata.no (1992)
44. Kuo, C., Romanosky, S., Cranor, L.: Human selection of mnemonic phrase-based passwords. In: Proceedings of the Second symposium on Usable privacy and security, *ACM International Conference Proceeding Series*, vol. 149, pp. 67–78. ACM Press (2006)
45. Lee, C., Lee, H.: A Password Stretching Method Using User Specific Salts. In: Proceedings of the 16th international conference on World Wide Web, pp. 1215 – 1216 (2007)
46. Leveugle, R.: Early Analysis of Fault-Based Attack Effects in Secure Circuits. IEEE Transactions on Computers **56**(10), 1431–434 (2007)
47. Li, X.Y., Teng, S.H.: Practical Human-Machine Identification over Insecure Channels. Journal of Combinatorial Optimization **3**(4), 347–361 (1999)
48. Lillesund, M.: Synlig login skaper utrygghet på Legevakten. www.idg.no/nyheter/article96717.ece (2008)
49. Lundh, F.: Slik oppbevarer Norges storste sykehus topphemmelig pasientinformasjon. www.vg.no/helse/artikkel.php?artid=578335 (2009)
50. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of Fingerprint Recognition, 2nd edn. Springer Professional Computing. Springer (2003)
51. Mansfield, T., Kelly, G., Chandler, D., Kane, J.: Biometric Product Testing Final Report. CESG/BWG Biometric Test Programme Issue 1.0, Centre for Mathematics and Scientific Computing, National Physical Laboratory (2001)
52. Martin, A., Phillips, P., Przybocki, M., Wilson, C.: An introduction evaluating biometric systems. Computer **33**(2), 56 – 63 (2000)
53. Matsumoto, T.: Human-computer cryptography: An attempt. In: Proceedings of the 3rd ACM conference on Computer and communications security, pp. 68–75 (1996)
54. Murray, M.: Gait as a total pattern of movement. American Journal of Physical Medicine **46(1)**, 290–332 (1967)
55. Narayanan, A., Shmatikov, V.: Fast dictionary attacks on passwords using time-space tradeoff. In: Proceedings of the 12th ACM conference on Computer and communications security, pp. 364–372 (2005)
56. NIST: The 2008 NIST Speaker Recognition Evaluation Results. www.itl.nist.gov/iad/mig/tests/sre/2008/official_results/index.html (2008)
57. Norsk Telegrambyrå: Datakollaps slettet 100.000 helsemeldinger. www.digi.no/php/art.php?id=787758 (2008)
58. Objectif Sécurité: ophcrack. ophcrack.sourceforge.net. Last visited: 22.3.2010
59. Oechslin, P.: Making a Faster Cryptanalytic Time-Memory Trade-Off , *LNCS*, vol. 2729, chap. Advances in Cryptology - CRYPTO 2003, pp. 617–630. Springer Berlin / Heidelberg (2003)
60. O'Gorman, L.: Comparing passwords, tokens, and biometrics for user authentication. Proceedings of the IEEE **Vol. 91, Issue 12**, 2019–2040 (2003)
61. Phillips, P., Flynn, P., Scruggs, W., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the Face Recognition Grand Challenge. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), pp. 947–954 (2005)
62. Phillips, P., Grother, P., Micheals, R., Blackburn, D., Tabassi, E., Bone, M.: Facial Recognition Vendor Test 2002 -Evaluation Report. Tech. rep., NIST (2003)
63. Phillips, P., Moon, H., Rizvi, S.: The FERET Verification Testing Protocol for Face Recognition Algorithms. Tech. Rep. NISTIR 6281, NIST (1998)

64. Phillips, P., Moon, H., Rizvi, S., Rauss, P.: The FERET Evaluation Methodology for Face-Recognition Algorithms. Tech. Rep. NISTIR 6264, NIST (1998)
65. Phillips, P., Scruggs, W., O'Toole, A., Flynn, P., Bowyer, K., Schott, C., Sharpe, M.: FRVT2006 and ICE2006 Large-Scale Results. Tech. rep., NIST (2007)
66. Phillips, P.J., Scruggs, W.T., OToole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M.: FRVT 2006 and ICE 2006 Large-Scale Results. NISTIR 7408, NIST-National Institute of Standards and Technology (2007)
67. Pinkas, B., Sander, T.: Securing passwords against dictionary attacks. In: Proceedings of the 9th ACM conference on Computer and Communications Security, pp. 161–170. ACM Press (2002)
68. Project-RainbowCrack: RainbowCrack. `project-rainbowcrack.com/index.htm`. Last visited: 22.3.2010
69. Ross, A., Jain, A.: Biometric sensor interoperability: a case study in fingerprints. In: Proceedings of BioAW 2004, LNCS 3087, pp. 134–145 (2004)
70. Sandström, M.: Liveness Detection in Fingerprint Recognition Systems. Master's thesis, Linkping University, Department of Electrical Engineering (2004)
71. Sasse, M., Brostoff, S., Weirich, D.: Transforming the "weakest link" - human/computer interaction approach to usable and effective security. BT Technol **19**(19), 122–131 (2001)
72. Sato, H.: N $\pm\varepsilon$: Reflecting Local Risk Assessment in LoA, chap. On the Move to Meaningful Internet Systems: OTM 2009, pp. 833–847. Springer-Verlag Berlin Heidelberg (2009)
73. Schmidt, J.M., Kim, C.H.: A Probing Attack on AES, chap. in Information Security Application, pp. 256–265. Springer Berlin / Heidelberg (2009)
74. Schreurs, N.: Snakker til PC og pasient. `www.idg.no/nyheter/article90664.ece` (2008)
75. Smith, R.: Authentication from passwords to public keys. Addison Wesley (2002)
76. Sosial- og Helsedirektoratet (utgivere): Norm for informasjonssikkerhet i helsesektoren. `www.helsedirektoratet.no/normen` (2006)
77. Stang, Øyvind., Snekkenes, E.: Experimental security evaluation of correlation based gait authentication. In: Proceedings of the 12th Nordic Workshop on Secure IT Systems (2007)
78. Tamviruzzaman, M., Ahamed, S.I., Hasan, C.S., O'brien, C.: ePet: when cellular phone learns to recognize its owner. In: Proceedings of the 2nd ACM workshop on Assurable and usable security configuration, pp. 13–18 (2009)
79. Unanue-Zahl, P.: Husnøkler kan kopieres fra bilder. `www.vg.no/teknologi/artikkel.php?artid=523553` (2008)
80. Watson, C., Wilson, C., Indovina, M., Cochran, B.: Two Finger Matching With Vendor SDK Matchers. NISTIR 7249, NIST -National Institute of Standards and Technology (2005)
81. Weinshall, D.: Cognitive Authentication Schemes Safe Against Spyware (Short Paper). In: Proceedings of the 2006 IEEE Symposium on Security and Privacy (S&P'06), pp. 295–300 (2006)
82. Woodward Jr., J.D., Orlans, N.M., Higgings, P.T.: Biometrics. McGraw-Hill/Osborne, California, USA (2003)
83. WP4: D4.11: eHealth identity management in several types of welfare states in Europe. `www.fidis.net/resources/deliverables/interoperability/d4-ehealt-identity-management-in-several-types-of-welfare-states-in-europe` (2008)
84. Yan, J., Blackwell, A., Anderson, R., Grant, A.: Password memorability and security: Empirical results. Security & Privacy, IEEE **2, Issue 5**, 25 – 31 (2004)
85. Zachariassen, E.: Ny lov åpner for misbruk. `http://www.tu.no/it/article212867.ece` (2009)
86. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face Recognition: A Literature Survey. ACM Computing Surveys **35**(4), 399–458 (2004)
87. Zhuang, L., Zhou, F., Tygar, J.: Keyboard acoustic emanations revisited. In: Proceedings of the 12th ACM conference on Computer and communications security, pp. 373–382 (2005)

Part II
Surveys in the Health Services

Chapter 7
Security Models for Electronic Medical Records

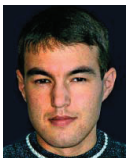
Davrondzhon Gafurov, Kirsi Helkala, and Nils Kalstad Svendsen

In **Security Models for Electronic Medical Records**. *Telektronikk*, 101 (1), 2005

Security models for electronic medical record

DAVRONDZHON GAFUROV, KIRSI HELKALA AND NILS KALSTAD SVENDSEN

In this article we give a definition of the electronic medical record (EMR) and summarize the main legal matters related to it. Further, we propose an information flow model of the current Norwegian EMR. This model is used to show that current access rights to the EMR are too general, and that requirements on accountability are not fulfilled. Finally we show how the use of dynamic access rights and non-repudiation can be used to improve the current situation.



Davrondzhon Gafurov is a PhD student at Gjøvik University College

Kirsi Helkala is a PhD student at Gjøvik University College

Nils Kalstad Svendsen is a PhD student at Gjøvik University College

1 Introduction

Over the last ten years our society has undergone an electronic revolution, and literally no segment of society in the industrialized countries is left untouched by the possibilities given by a multitude of computer applications and network services. The health sector is no exception, however the following characteristics differentiate this sector from the others:

1 The sensitivity of information treated in the system, emphasized by strict legal restrictions;

2 The complexity due to the wide range of systems and the large number of users;

3 The tight connection between private and public health sector.

Based on a general definition of the medical record and on the legal framework, we seek a classification of information sharing among health personnel working in the same organization, and communication between personnel from different organizations. Further, we compare the Norwegian model with the British Medical Association model (see Anderson [4]). As an extension of this model, we show how to model the EMR as a database with dynamic access rights, such that the requirements of confidentiality, integrity, availability and accountability are met. Finally we give an example of how the above framework can be used to achieve non-repudiation in message exchanges among health personnel.

2 What is an electronic medical record?

Based on the Health Personnel Act [12] and the Patients' Rights Act [14], KITH[1)] defines an electronic patient record as: "An electronic collection of registered information on a patient related to health care". We find this definition and the term "electronic patient record" limiting. We consider Lærum's following discussion on the term electronic medical record (EMR) in [16] more appropriate in this context:

> *The EMR in its simplest form may be regarded as an electronic version of the paper-based medical record. It is the repository of clinical information on which health personnel base their decisions regarding health care of the individual patient. However, its content is not universally defined in the literature, and consequently, the concept is named in a multitude of ways.*

Lærum gives an overview of different definitions of EMR. For our purpose, to give general considerations of how to manage access rights to the database to ensure that requirements to confidentiality, integrity, availability and accountability are met, the general definition from [17] is appealing. Here the EMR is defined as a database containing data from various sources as shown in Figure 1. To our knowledge, a system like this is currently not implemented in any Norwegian hospital. Most hospitals use DIPS[2)],

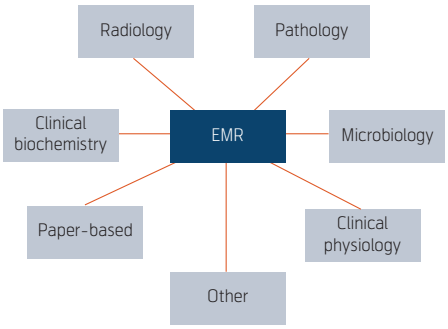


Figure 1 The EMR database

1) *Kompetansesenter for IT i Helse- og sosialsektoren AS (Norwegian Centre for Informatics in Health and Social Care)*
2) *www.dips.no*

DocuLive EPR[3] or InfoMedix[4] to process health data. The solutions from these vendors are especially adapted to each institution to fit into the local systems. As academics, we are not constrained by these practical obligations and have the liberty to work with an idealized scenario. We propose an information theoretic model of how access to the EMR database should be administrated. But first we give a brief summary of the legal framework.

3 Legal restrictions

The following Norwegian legislation regulates the EPR: The Health Personnel Act [12], the Patients' Rights Act [14], Personal Health Data Filing System Act [13], the Journal Act [7], Personal data regulations [11] and the Archives Act [10]. The main features in these laws related to information security are:

1 Only the data controller, the data processors and people working under the instruction of the controller or the processor may be granted access to personal health data. Access may only be granted if this is necessary for the work of the person concerned and in accordance with the rules that apply regarding the duty of secrecy.

2 The health care provider is obliged to enter or record information in a patient record for the individual patient. The duty to keep patient records does not apply to co-operating personnel providing care in accordance with instructions or guidance from other health personnel.

3 Health institutions are obliged to designate one person with superior responsibility for the individual patient record including making decisions relating to what information is to be entered into the patient record.

4 It must be evident from the records who has entered the information into the patient records.

5 Corrections must be carried out through re-entering the information of the patient records, or by adding a dated correction in the records. Corrections must not be made by deleting information or comments.

6 Upon demand from the person whom the information in the patient record relates to, or of their own accord, health personnel can under certain conditions delete information or comments in the patient record.

7 Patient records may be kept electronically.

8 The data controller and the data processor must by means of planned, systematic measures ensure satisfactory data security with regard to confidentiality, integrity, quality and accessibility in connection with the processing of personal health data.

9 In legal or administrative cases relating to the professional conduct of health personnel, notes recorded in patient records, patient records and patient record material may be required for the purpose of being presented as evidence, either as originals, or as certified photocopies or printouts.

10 The patient is entitled to access to his or her medical records with enclosures and upon special request is entitled to a copy. Under certain circumstances the patient may be denied access to information in his or her medical records.

11 The patient is entitled to object to the disclosure of his or her medical records or information in the records. Furthermore, the information may not be disclosed if there is reason to believe that the patient would have objected to this if asked.

This legal framework poses strong requirements on availability, confidentiality, integrity and accountability, and states that patient approval is often necessary. Our belief is that an implementation of dynamic access rights and PKI is necessary to fulfill these requirements. We also note that the most common operations are read and write. Deletion and overwriting are relatively rare operations, and may be avoided by adding a deletion tag to an element instead of deleting it. In this report, we therefore focus on the management of read and write access rights.

4 Classification

The KITH report [2], which gives an overview of the need for PKI in the health sector, divides the PKI applications in the health sector into the following three groups:

- Communication within public administration/ health service;
- Communication between public administration/ health service and clients;

[3] *www2.siemens.no/med/*
[4] *www.infomedica.no*

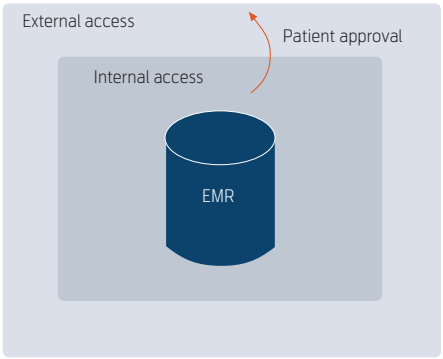


Figure 2 Simplistic view of ownership and access right to the EMR

- Communication between public administration and industry.

This focus is slightly different from ours; as we are concerned about who owns a medical record, who does not own a medical record, and how do these communicate between each other. This point of view is based on the first legal aspect of Section 3. Internally, the local administrator grants access rights, whereas access for external health personnel requires patients' agreement. Figure 2 illustrates the separation between these two cases.

4.1 Internal access policy

Internal information exchange takes place inside an organization. The organization has one manager who is responsible for all the employees. An organization might for example be the practice of a general practitioner, a hospital or united hospitals. From a legal aspect, the local data controller is the one who internally grants access to the data and ensures satisfactory information security with regard to confidentiality, availability, integrity and quality. A description of how this is done today in one of the Norwegian health regions is given in Section 5.

4.2 External access policy

External information exchange takes place when health information is transferred between working units, which have different managers. Scenarios of these situations are communication between a general practitioner and a specialist in a hospital or between a private laboratory and a general practitioner.

While the KITH report [2] focuses on all communications with a need for PKI within the health sector, we have tried to identify the communications where direct access to the EMR is necessary for at least one of the involved parties. These scenarios are summarized in Table 1. In Figure 3 we show the parties involved in this communication and we emphasize those who are in possession of an EMR. In Section 6 we treat the problem of reading and writing information to the different parts of the EMR without violating the legal framework. We especially focus on the case where laboratory results are sent from the laboratory to a doctor, and how the laboratory can be in the position to prove that the doctor was in possession of the result at a given time.

5 Models of Norwegian health service

Having classified different communication categories, this section introduces models of the information flow in Norwegian health service. The internal access model is based on the idea that there are two layers in every health institution. The outer layer consists of different departments and the sub-model

Communication	Authorization	Non-repudiation	No. of actors involved
Referral	X	X	2'
Case summary	X	X	2
Prescription	X	X	3
Journal transfer	X	X	2'
Lab answers		X	2
Sick leave		X	3
Medical certificate		X	3

Table 1 External communications where at least one of the parties needs direct access to EMR. This table is an abstract of the table given in the KITH report [2]. All communications require authentication, encryption and integrity. Non-repudiation for lab answers is not in the KITH report, but has been added to facilitate the use of the EMR as legal evidence. When the number of actors involved is marked with " ' " it means that patient approval could be included in the communication

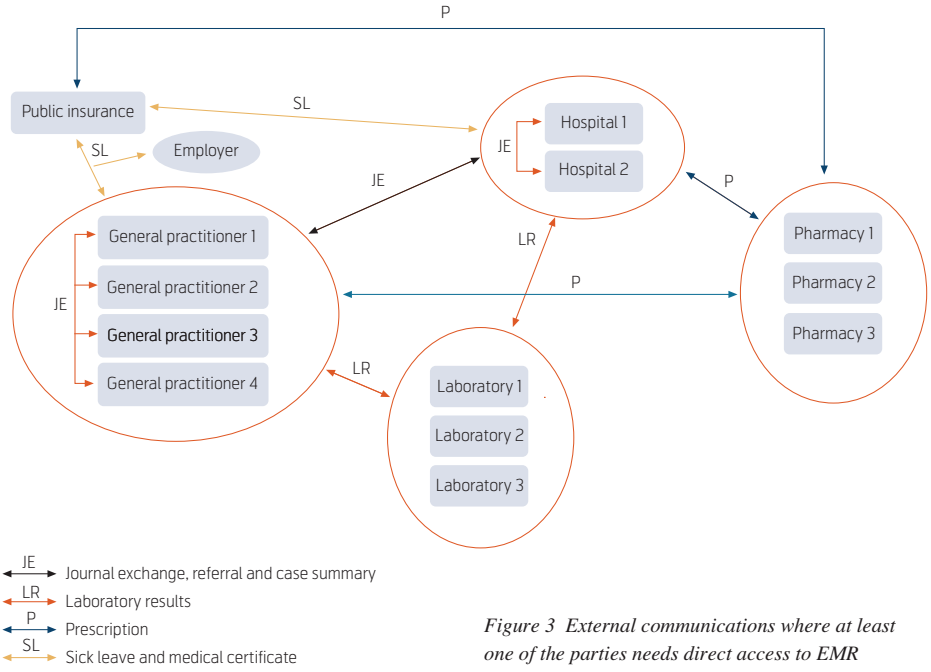


Figure 3 External communications where at least one of the parties needs direct access to EMR

explains the information flow between the departments. The inner layer consists of only one department and a sub-model is created to describe the information flow within the department. The majority external information flow is still based on the exchange of paper documents and use of fax machines.

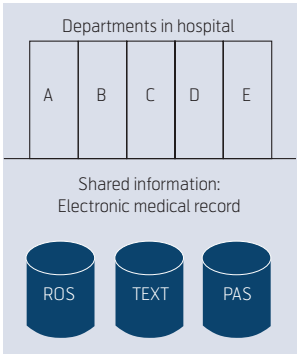


Figure 4 Multilateral policy in hospitals. ROS = Result of samples, TEXT = Text part of the patient record, often called the journal, PAS = patient administration system. Note: The EMR can contain more elements than shown in the figure

5.1 Internal access policy

A hospital can be divided organizationally into different departments as shown in Figure 4. This model can be generalized. If we were modeling a clinic with one doctor the model would contain only one department. Figure 4 also shows how different departments have access to the same EMR. This is the outer layer of the information flow, and can be modeled by a multilateral security policy.

BMA security policy. Multilateral security policy, shown in Figure 5, prevents information flowing across departments. One of the multilateral security policies is The British Medical Association security

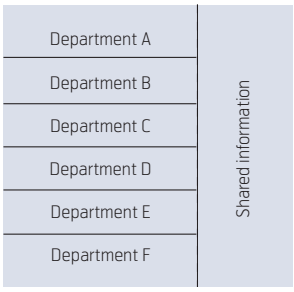


Figure 5 Multilateral security policy

policy BMA proposed by Anderson [3]. It was designed specifically for the needs of the health service. BMA security policy consists of nine principal elements: access control, record opening, control, consent and notification, persistence, attribution, information flow, aggregation control, and trusted computing base.

A comparison of the information given by [3, 9, 15] shows that Norwegian electronic patient record system contains most of the principal elements of the BMA model, and can therefore be considered as similar to the BMA model.

Access control:
- In the BMA model, records are marked with an access control list naming the people or groups who may read them and append data to them.

- The Norwegian system is similar because health workers are divided into certain user groups based on their position in the hospital and these groups have different access rights to the records.

Record opening:
- In the BMA model, the patient might have multiple records with different sensitivity level. A clinician may open a new record with the patient. Where the patient is referred, the clinician may open a record with the patient and the referring clinicians.

- In the Norwegian system, records can be opened by anyone who has the right and obligation to do so. Patients are allowed to see their record in the presence of doctors.

Control:
- In the BMA model, one of the authorized clinicians may alter the access control list and add the health care professionals to it.

- In the Norwegian system, access privileges are given by the IT department or administrative department based on the recommendation given by the system owner. In the case of the internal information exchange the system owner is the manager of the united hospitals.

Consent and notification:
- In the BMA model, the responsible clinician must notify the patient whenever his patient record's access control list has been altered.

- In Norway, patients' permission will be asked when patient records are sent from one clinic to another, but inside the hospital this is not done. The reason for this is that a hospital is considered to be one clinic where all the workers already have previously mentioned rights to see the records.

Persistence:
- In the BMA model, no one should be able to delete the patient records before the expiration date.

- In the Norwegian system, normal users are not allowed to delete records and even deleting a sentence in the records is denied. Users are able to correct information by adding new information. If deleting is considered to be necessary then there is a certain protocol to follow. Request has to be made for the system owner and depending on the answer the IT department can perform the deletion.

Attribution:
- In the BMA model, all accesses to patient records must be marked on the record.

- The Norwegian system is similar because every access to the patient records will be logged (username and time, no digital signature on added information).

Information flow:
- In the BMA model, information from a less sensitive record can be added to a more sensitive record, but not the other way round.

- This is also true in Norway, but it is not relevant in our model because we consider only one medical record at a time.

Aggregation control:
- In the BMA model, patients must receive special notification if a person who has access to patient records on a large number of people, is proposed to be added to their access control list.

- There is no such requirement in the Norwegian laws.

Trusted computing base:
- In the BMA model, computer systems that handle personal health information must have a subsystem that enforces the above principals in an effective way.

- In Norwegian hospitals, there are three different previously mentioned EMR applications, which are based mostly on these principals.

Inside the hospital wards, the policy is not multilateral, but multilevel, because people working in the hospital, even in the same ward, have different clearance levels. These clearance levels, and therefore the

access rights to the patient records depend on a user's position and working place in the hospital. Doctors have read and write access, nurses have read and limited write access and the secretaries in the wards have limited read and write access to the records. This indicates that the information flow inside the wards can be modeled by using the Bell-La Padula model [4].

Bell-La Padula confidentiality model. According to [4], the Bell-La Padula model belongs to the class of multilevel security policies meaning that it handles data at different sensitivity levels and prevents information flowing down in the hierarchy. Documents and users are given clearance levels. Users are able to read information written by users having the same or a lower clearance level than themselves and write to documents written by users having the same or superior clearance levels. The only one who can write down is the one who has no access to upper clearance level data. These principles are shown in Figure 6.

Some examples of the current situation inside the wards are given in Figures 7, 8 and 9. Figure 7 shows the information flow in a typical clinical ward. Nurses have limited write access to the medical record's text part (mentioned in the figure as NTEXT), where they write down medication and treatment they have given to the patient. However they are not allowed to see the whole text journal. As shown in Figure 7, doctors currently have this right. Figures 8 and 9 show information flows from radiography and laboratory departments. The full model of internal information exchange consists of two previously mentioned sub-models, Bell-La Padula and BMA.

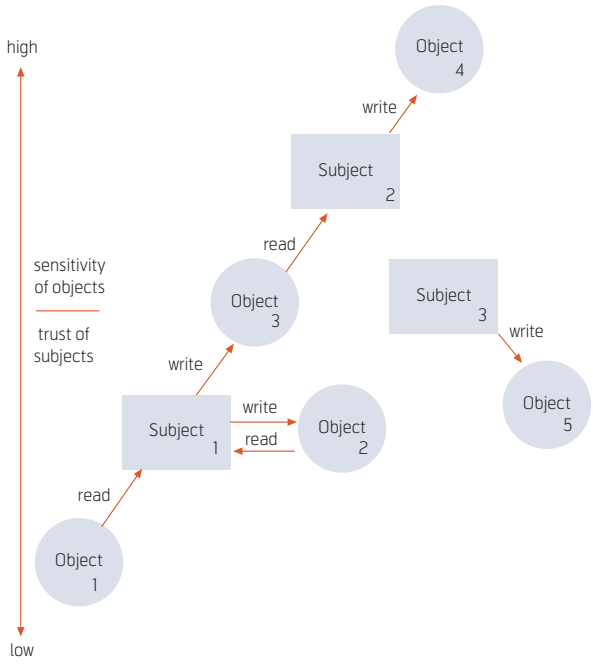


Figure 6 Bell-La Padula Confidentiality Model

5.2 External access policy

Today's situation is more or less paper based information exchange. For example, all medical prescriptions are written on a piece of paper, which is taken to the pharmacy by the patient. General practitioners have the possibility to send and receive referrals, case summaries and laboratory results by using electronic

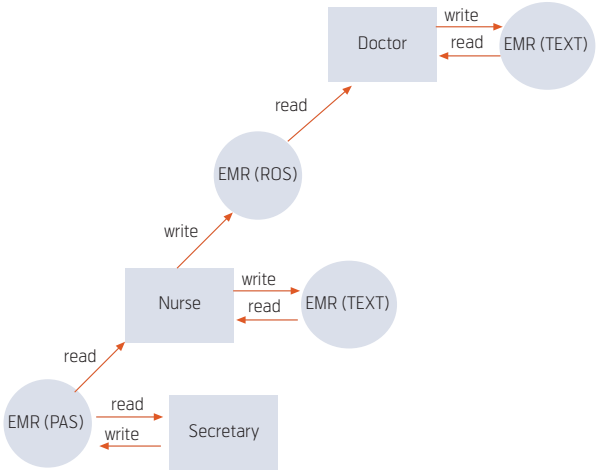


Figure 7 Information flow in the clinical ward

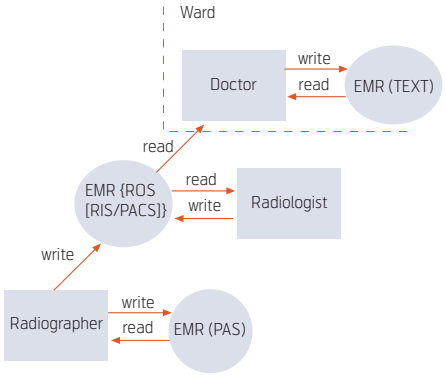


Figure 8 Information flow in the radiology department. RIS is the roentgen information system and PACS is the picture archive and communication system

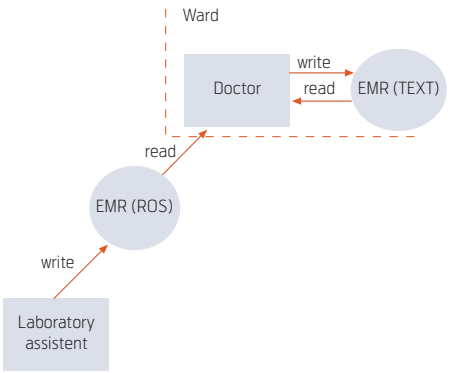


Figure 9 Laboratory department

secure envelope, which uses xml-protocol, but not all of them do so. Even if the document is in electronic form, it has to be stored in the health record by the general practitioner him- or herself. If the document is in paper form, the storing is time consuming because the practitioner has to scan it to the records. According to the Health Personnel Act (see Section 3) the private service providers or the hospitals do not have access rights to the medical records kept by the general practitioner, or vice versa. If the whole medical record needs to be transferred from one place to another, permission has to be asked from the patient and the medical record is sent as a paper document after receiving permission.

5.3 Weaknesses in the current system

Our models show that doctors working in the same department have the same access rights to patient records. This means that a doctor can access record information on a patient without being directly involved in the treatment of the patient. This is a violation of the first point in the list of legal restrictions stating that access to personal health data should only be granted if it is necessary for the work of the person concerned.

Access to the EMR is logged, but there is no use of digital signatures. This is a weakness if the EMR is to be used as evidences in a legal case as it can be claimed that the password had been stolen, the computer had been left open and so on. The use of a digital signature would make such statements less plausible.

Finally the external information exchange might be the greatest weakness of the system. The use of post or a courier to transfer paper files is an outdated method. This can be done quicker and safer by the use of a non-repudiation protocol to negotiate and carry out the exchange.

Motivated by the observations above, in Section 6 we propose models for administrating the access rights to the EMR such that:

1 Access to the EMR is more restricted than before and limited to only personnel currently working with a patient.

2 Accountability of actions is assured to a larger extent than before.

The proposed solutions are based on the existence of an internal and external PKI framework.

6 Models for the future

As pointed out, one may have the impression that compared to the legal constraints on patient data, access to EMR is too general and that the requirement of accountability may not be fulfilled. In this section we show how role-based access control (RBAC) and non-repudiation can be used to improve this situation.

6.1 Role-based internal information exchange

Motivation. The EMR is distributed among various places, e.g. laboratories and hospitals, and accessed by different users, e.g. doctors and nurses. The users of the system possess different access policies. Moreover, the nature of these access policies is dynamic. Dynamic access policy means that granting of operation on objects is done based not only on user's predefined set of functionality, but also based on some other contextual information (e.g. time or location). For example, a doctor in an operating theatre should have full access to patients' records but on other hand, from his home PC he should have limited access to patients' EMR.

A role-based access control is a promising access policy that can address dynamic policies. A core feature of the RBAC is that permissions (or rights or access rights) are assigned to roles, and users are members of the appropriate roles. Moreover, RBAC supports three well-known security principles: least privileges, separation of duties, and data abstraction [18]. Assignment of permissions to roles rather than to users seems a more natural way from the organizational point of view because roles are associated with work positions and duties of users within the organization. A study [6] shows that permissions assigned to roles tend to change relatively slowly compared to changes in the user membership of roles.

Role-based access control model. RBAC is defined in terms of four model components: Core RBAC, Hierarchical RBAC, Static Separation of Duty Relations

and Dynamic Separation of Duty Relations [5]. We consider only a core RBAC with static and dynamic separation of duties. A core RBAC reference model proposed by NIST is depicted in Figure 10. The elements in the RBAC model are the set of users (*U*), roles (*R*), permissions (*P*) and their relationships that are user-to-role assignment (*UA*) and role-to-permission assignment (*PA*). A *role* is a job position or work title associated with a set of functions or responsibilities within the organization, for example nurse, laboratory assistant, etc. A *permission* is a predefined type of *operation* on an *object* within the system, for example permission to read an X-ray of a patient where read is an operation and *X-ray* is an object. A *session* is a mapping of a user to the activated subset of the roles assigned to the user. The functions *session_roles* and *user_session* returns the roles activated by session and the set of sessions associated with a user, respectively. The user assignment and permission assignment are many-to-many relationships. That is, one user can be assigned to several roles and one role to several users; and one role can be granted several permissions and one permission can be granted to several roles.

An RBAC with static separation of duties and dynamic separation of duties is an extension of the core RBAC model, see Figure 11. Static separation of duties (SSD) places constraints on the assignment of users to roles. Dynamic separation of duties (DSD) specifies constraints on permissions of the user to activate assigned roles. From another point of view, the difference between SSD and DSD is that SSD is set during design time while DSD is set during runtime.

RBAC model for EMR. We can show an example of RBAC for a model of internal information access being described in Section 5.1. We have the following sets:

- *U* = {*User_1, User_2, …, User_n*};
- *R* = {*Doctor, Nurse, Secretary, Radiologist, Radiographer, LaboratoryAssistant, Patient*};
- *P* = {*ReadText*[5], *WriteText, ReadROS*[6], *WriteROS*, *ReadPAS*[7], *Write-PAS*, *ReadRIS*[8]/*PACS*[9], *WriteRIS*/*PACS*}.

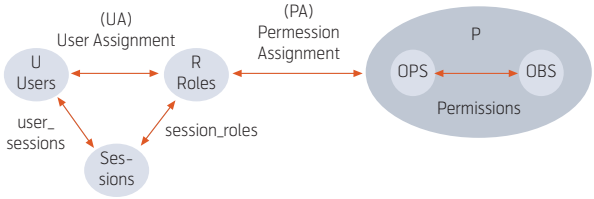


Figure 10 Core RBAC

$User_i$ can be named by using SSN[10] or some other identifier of the particular doctor, nurse and other users of the system[11]. And also, instead of dividing objects and operations on them, we have defined them together as a set of permissions *P*. We define user assignment as the function *user_roles*, i.e. a set of roles for the given user.

- *user_roles*(*User_1*) = {*Doctor*}
- *user_roles*(*User_2*) = {*Nurse*}
- *user_roles*(*User_3*) = {*Secretary*}
- *user_roles*(*User_4*) = {*Radiologist*}
- *user_roles*(*User_5*) = {*Radiographer*}
- *user_roles*(*User_6*) = {*LaboratoryAssistant*}
- *user_roles*(*User_7*) = {*Patient*}

We also define permission assignment as a function *role_permissions* that gives us a set of permissions for the given role:

- *role_permissions*(*Doctor*) = {*ReadText,WriteText,ReadROS,ReadRIS/PACS*}

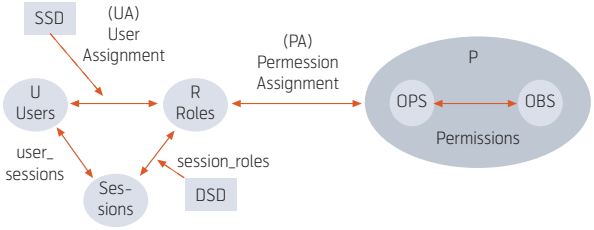


Figure 11 RBAC with static and dynamic separation of duties

5) *Text is the text part of the patient record*
6) *ROS is result of samples*
7) *PAS is the patient administration system*
8) *RIS is the roentgen information system*
9) *PACS is the picture archive and communication system*
10) *Social security number*
11) *Such as the identifyer in the Norwegian health personnel register*

- *role_permissions*(*Nurse*) = {*ReadPAS, WriteROS*}
- *role_permissions*(*Secretary*) = {*ReadPAS, WritePAS*}
- *role_permissions*(*Radiologist*) = {*ReadRIS/PACS, WriteRIS/PACS*}
- *role_permissions*(*Radiographer*) = {*WriteRIS/PACS, ReadPAS, Write-PAS*}
- *role_permissions*(*LaboratoryAssitent*) = {*WriteROS*}

As mentioned earlier, rules regulating SSD can be set during design time, and constraints on user's role assignments can be specified. For example, a user cannot be assigned to both of the roles *Doctor* and *Pharmacist*. On the other hand, DSD puts constraints on activating assigned roles. For example, *Radiographer*'s permission to write to EPR (RIS/PACS) is activated when the doctor requests and sends the patient to take X-ray. Until that time a radiographer is not allowed to write RIS/PACS although he has such a right. Thus, for our purpose we define dynamic access policy as granting permission to a role based on some available runtime information. To address the dynamism we define the following abstract function, which serves as a precondition for activating roles:

can_activate(*r, p, t*)

This function returns true if the user with the role *r* can perform permission *p* based on some information available at time *t*, and false otherwise. In fact, the parameter *t* may encapsulate not only time, but other information like location as well. In [8], a spatial role-based access control framework applied to health care is presented. The framework utilizes *location* information in access control decisions, in order to determine the permissions a role encompasses at given location.

Accountability RBAC. In order to be compatible with governing rules and laws in our model, we assume that a local Trusted Third Party[12)] (TTP) is running. The aim of TTP is to collect evidence of the users' activities by recording their operation on EMR. In its simplest form, the TTP can store data in log files.

6.2 External information exchange

Motivation. The external exchanges shown in Table 1 can be divided into two cases; communication between two main actors or communication among three main actors. Sending case summaries, referrals, records and laboratory results belongs to the model where there are only two main actors. The rest, prescription, sick leaves and medical certificates belong to the model where there are three main actors. These are shown in Figure 3. In this section we focus on how non-repudiation can be used to improve the accountability of these communications. We first give a brief introduction to non-repudiation.

Non-repudiation. Anyone slightly familiar with information security can list four of the main categories of security services defined by ISO 7498-2 [1]: Authentication, access control, confidentiality and data integrity. Fewer are aware of the fifth category, namely non-repudiation. Non-repudiation is related to authentication and data integrity, but has stronger proof requirements and is a protection against false denial of having been involved in a communication. The general goal of a non-repudiation service is to collect, maintain, make available, and validate irrefutable evidences concerning a claimed event or action and to resolve disputes about the occurrence or the non-occurrence of that event or action. Typical conflicts that can be solved by non-repudiation are the following:

- A claims having sent M to B while B denies having received M.
- B claims having received M from A while A denies having sent M.
- A claims having sent M before time T while B denies receiving it before T.

The main idea is that the parties in a communication should be in possession of a receipt when the communication is terminated. This receipt should be generated in such a way that both parties have the same opportunity to cheat, meaning that even though the computational power of one of the parties is superior to the other party, it should not have better possibilities of creating false evidence.

There are usually three parties involved in a transaction assuring non-repudiation: Originator, recipient and trusted third party (TTP). In a non-repudiation protocol, the TTP can have different degrees of involvement in the communication (inline, online or offline) and has to play different roles (certification authority, notary, delivery authority, time stamping authority and adjunctor). For more on non-repudiation and the role of TTPs see [19]. The evidences,

12) *The role of the TTP is defined in Section 6.2.*

or receipts, can mainly be generated by two types of security mechanisms:

- Secure envelopes generated by TTPs using symmetric cryptographic techniques;
- Digital signatures generated by any party using asymmetric cryptographic techniques.

The use of secure envelopes requires an unconditional trust of the TTP, which in many cases is unrealistic. Therefore many users prefer digital signature based on evidences where the trust in the TTP as a certification authority can be reduced by appropriate registration and certification procedures.

Example. Let us now take a look at the two actors model and consider a communication between a general practitioner's (GP's) office and a laboratory with the purpose to analyze a patient's sample. In a real life situation, the main part of the communication is handled by the secretaries or assistants, as Figure 12 shows. Our model takes this into account. Our model also obeys the following privileges given to different roles: nurses and secretaries have limited access rights to the EMR, while the doctors have full reading and writing access. Limited access gives full rights to handle patient administration data but not the medical data. Results of the samples and diagnoses of the sickness are private issue between doctor and the patient.

The non-repudiation protocols between actors can be used for achieving confidentiality and integrity of the information exchange. Internal PKI can be implemented in the laboratory and at the doctor's office to achieve non-repudiation in that part of the communication too, but is not treated in this article.

Protocol. Before stating the protocol for the exchange, we introduce the necessary general notation.

- A, B and U: actors of the communication. U can be both A or B.
- M: message sent from A to B.
- K: message key defined by A.
- C = eK(M): commitment (cipher text) for message M.
- L = H(M,K): a unique label linking C and K.
- f_i (i = 1, 2, ...): flags indicating the intended purpose of a signed message.

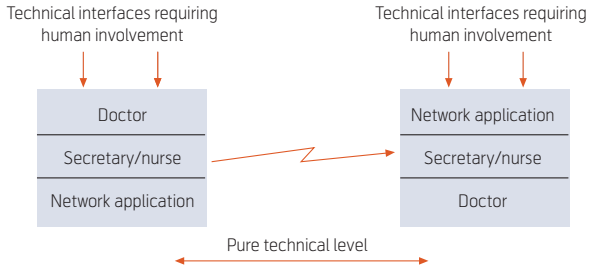


Figure 12 Three layers communication

- sub_K = sS_A (f_5, B, L, K, H(C)): authenticator of K provided by A.
- con_K = sS_{TTP} (f_6, A, B, L, K): evidence of confirmation of K issued by the TTP.
- abort = sS_{TTP} (f_8, A, B, L): evidence of abortion of a transaction issued by the TTP.
- T_{sub} is the deadline that A should either send a message key or submit it to the TTP.

The first part of the protocol handles booking of an appointment at a laboratory. The general practitioner's (GP's) secretary (GPS) sends a request to the laboratory. The receiver in the laboratory is also a secretary (LABS) who handles the first part of the communication. Based on the NR2 protocol described in [19], the communication goes as in Figure 13. The message contains the patient information and a description of what the laboratory should do. After receiving the final answer from the laboratory, the patient or the sample can be sent for testing. Using for example a courier to send the sample will ensure non-repudiation in this phase. Conflict resolution for this protocol is described in [19].

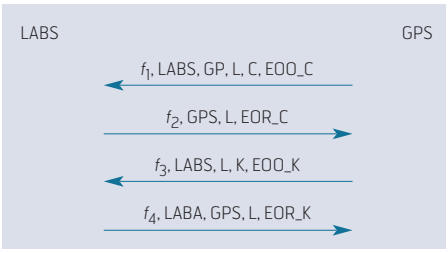


Figure 13 The NR2 protocol. Where EOO_C = $sS_{GPS}(f_1, LABS, GP, L, C)$, EOR C = $sS_{LABS}(f_2, GPS, L, C, T_{sub})$, EOO_K = $sS_{GPS}(f_3, LABS, L, K)$ and EOR_K = $sS_{GPS}(f_4, LABA, GPS, L, K)$

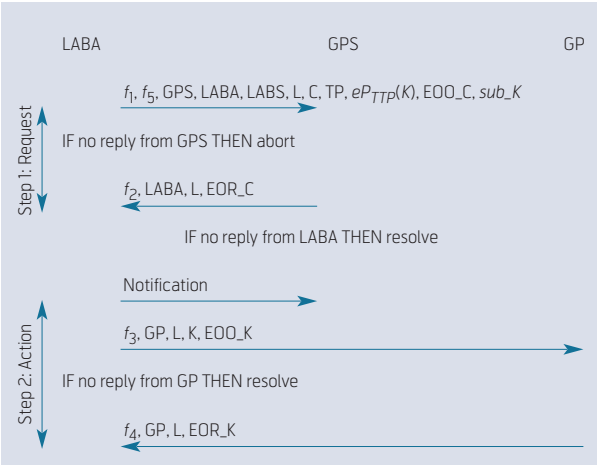


Figure 14 The main protocol, where EOO_C = $sS_{LABA}(f_1$, GP, GPS, LABA, LABS, L, C), EOR_C = $sS_{GPS}(f_2$, LABA, GP, L, C, $T_{sub})$, EOO_K = $sS_{LABA}(f_3$, GP, L, K) and EOR_K = sS_{GP} (f_4, LABA, L, K)

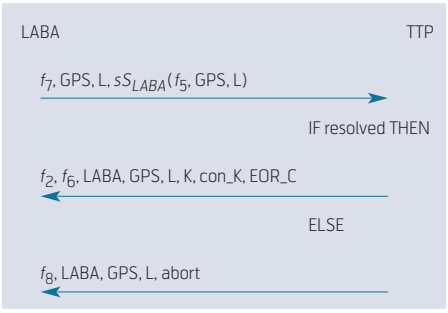


Figure 15 The abort subprotocol

When the analysis is ready, the second part of the protocol can be initiated. In this step, there are two main actors but three parties. The purpose of this is to separate the secretary's and the doctor's duties. The laboratory assistant sends the encrypted results to the GP's secretary using the protocol described in Figure 14 (which is based on the NR3 protocol described in [19]). After Step 1, the secretary has an encrypted laboratory result, which she can save to the EMR. The secretary cannot read the results because she does not have the key. This is sent to the doctor in the next step after the secretary has sent her confirmation back to the laboratory assistant.

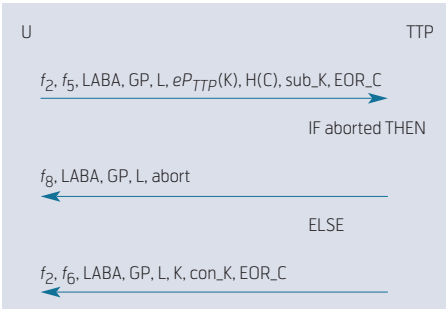


Figure 16 The resolve sub protocol

After having received the message from LABA in Step 2, the doctor has the key and can decrypt the result in the EMR. After receiving the final confirmation from the doctor, the laboratory assistant knows that the doctor has received the last result. The abort and resolve subprotocols are given in Figure 15 and Figure 16.

Problem places. In Step 1, the protocol will be aborted after the first message, if the laboratory does not receive the answer within a certain time period. Nobody has gained anything. The secretary then has four options:

1 The secretary saves the document and sends an answer: The protocol continues as normal and the doctor will receive the key eventually.

2 The secretary saves the document and does not answer: The protocol will be aborted by the laboratory. This protocol is described in Figure 15 and includes a TTP. The laboratory then has to send the first message again.

3 The secretary does not save the document but answers: The protocol will continue normally, but then the doctor has nothing to read later. In this case, the laboratory will be able to prove that the secretary has received the document. If the doctor locally does not trust his secretary to save it, a local PKI may be implemented to ensure non-repudiation.

4 The secretary does not save the document and does not answer: The protocol will be aborted and laboratory has to send the first message again.

If the GPS does not receive a notification from the LABA, the GPS can assume that the doctor has not received the decryption key and will, if time runs out, call for a resolution by using the protocol presented in Figure 16. It is therefore not in LABA's interest not to send the notification. On the other hand, the secretary can deny having received the notification even if LABA has sent it to her. This is the weakest point in our protocol. An easy, but relatively expensive, way to solve this problem is that the notification is sent through an online TTP. In this way the secretary cannot deny having received the notification.

After the message is sent from LABA to the GP, the doctor knows that the result has been sent earlier and he should be capable of retrieving the message and decrypting it in order to read the result. The doctor now has six options to proceed.

1 If the doctor finds the document from EMR, he reads it and answers: The laboratory will get the confirmation that the doctor has got the result.

2 If the doctor does not find the document from the EMR, he has to ask the secretary to save it to the EMR. He then proceeds as in case 1 and the laboratory will get the confirmation.

3 If the doctor does not find the document and the secretary cannot find it either, then the doctor knows that the secretary has made a mistake. To get the results, the communication has to be started again.

4 If the doctor does not read the document but answers, then the laboratory will get the confirmation that the doctor has the results.

5 If the doctor reads the document but does not answer, then the LABA calls for a resolution and will get evidence that it has sent the decryption key.

6 If the doctor does not read the document and does not answer, the LABA calls for a resolution and will get evidence that it has sent the decryption key.

In the case of a resolution the TTP generates a status report, gives evidences of what has happened and gives the message key by request. This protocol can be used for any information exchange between two main actors and their assistants. The protocol for the three main actors is to be defined in a future work. It can also be noted that in situations where a patient approval is necessary, the patient can be included as a party in a three actor protocol.

7 Conclusion

Based on a general definition of the electronic medical record as well as the legal framework for patient information, we have evaluated the current information flow in Norwegian health institutions. We have shown three examples of how the current methods do not meet the legal restrictions, and we have proposed a solution for two of them. Role based access control is proposed to limit the access to medical records in internal information exchange, and non-repudiation to achieve accountability for external information exchange.

Both solutions require the implementation of a public key infrastructure in the health sector. The proposed solution for external communication is an alternative to the propositions of a centralized health register. Our solution leads to effective communication among institutions, while avoiding the vulnerabilities introduced in collecting all the information in one database.

References

1 ISO. *Information processing systems – open systems interconnection – basic reference model – part 2: Security architecture*. International Organization for Standardization, 1989. ISO 7498-2.

2 Aksnes, B, Vestad, A, Henriksen, E, Skipnes, E, Kvaase, I E. *Forprosjekt for pki i helsenett, forprosjektrapport*. KITH, Sukkerhuset, 7489 Trondheim, January 2002. Technical Report R 3/02.

3 Anderson, R. A security policy model for clinical information systems. In: *IEEE symposium on Security and Privacy*, 1996.

4 Anderson, R. *Security engineering: A guide to building dependable distributed systems*. Wiley, 2001.

5 Ferraiolo, D, Sandhu, R, Gavrila, S, Chandramouli, R, Kuhn, D R. Proposed nist standard for role based access control. In: *ACM Transactions on Information and System Security*, 224–274, August 2001.

6 Ferraiolo, D F, Gilbert, D M, Lynch, N. An examination of federal and commercial access control policy needs. In: *Proceedings of the 16th NIST-NSA National Computer Security Conference*, September 1993.

7 *Forskrift om patientjournal*. Available from http://www.lovdata.no.

8 Hansen, F, Oleshchuk, V. Application of role-based access control in wireless healthcare information systems. In: *Scandinavian Conference in Health Informatics*, 30–33, June 2003.

9 Interview with J G Bronch, head of IT department in Sykehuset Innlandet 18.03.2005.

10 *Lov om arkiv (arkivloven)*. Available from http://www.lovdata.no.

11 *Lov om behandling av personopplysninger (personopplysningsloven).* Available from http://www.lovdata.no.

12 *Lov om helsepersonell m.v. (helsepersonelloven).* Available from http://www.lovdata.no.

13 *Lov om helseregistre og behandling av helseopplysninger (helseregisterloven).* Available from http://www.lovdata.no.

14 *Lov om pasientrettigheter (pasientrettighetsloven).* Available from http://www.lovdata.no.

15 Lærum, H. *What is an electronic medical record and how should it be evaluated?* Trondheim, NTNU.

16 Lærum, H. *Evaluation of electronic medical record – A clinical task perspective*. Norwegian University of Science and Technology, 2004. PhD thesis.

17 Nystadnes, T. *What is an electronic medical record and how should it be evaluated*. Dissertation lecture available at http://kvalis.ntnu.no/PublicDocs/2004-03-18-laerum-dissertation-lecture.ppt.

18 Sandhu, R S, Coyne, E J, Feinstein, H L, Youman, C E. Role-based access control models. In: *IEEE Computer*, 29 (2), 1996.

19 Zhou, J. *Non-repudiation in Electronic Commerce*. Computer Security Series. Artech House, first edition, 2001.

Davrondzhon Gafurov received his MSc in Computer Engineering from the Technological University of Tajikistan (TUT), Khujand, Tajikistan in 2000. From 2000 to 2004 he was Research Assistant at the Department of Programming and Information Technology at TUT working on NLP project. He is currently pursuing the PhD degree in information security at Gjøvik University College, Gjøvik, Norway. His current research interests focus on information security.

email: davrondzhon.gafurov@hig.no

Kirsi Helkala received her MSc in Mathematics from the University of Joensuu, Finland, in 2001. After graduating she worked as assistant in the University of Joensuu and then as a mathematics, physics, and chemistry teacher in Kesämäenrinteen Koulu, Lappeenranta, Finland. She is currently a PhD student at Gjøvik University College. Her work interest is personnel authentication and the title of her PhD project is "Authentication in a health service context".

email: kirsi.helkala@hig.no

For a presentation of Nils Kalstad Svendsen, please turn to page 64.

Chapter 8
Authentication in Norwegian Health Services

Kirsi Helkala

In *Proceedings of the International Symposium on Health Informatics and Bioinformatics*, Turkey '07, 2007

Part III
Use of Biometrics

Chapter 9

Gait Recognition Using Acceleration from MEMS

Davrondzhon Gafurov, Kirsi Helkala, and Torkjel Søndrol

In *Proceedings of The First International Conference on Availability, Reliability and Security (ARES 2006)*, pp. 432-439, 2006

Gait Recognition Using Acceleration from MEMS

Davrondzhon Gafurov, Kirsi Helkala, and Torkjel Søndrol
Norwegian Information Security Lab - NISlab
Department of Computer Science
and Media Technology
Gjøvik University College
P.O. Box 191, 2802 Gjøvik, NORWAY
{davrondzhon.gafurov, kirsi.helkala, torkjel.soendrol}@hig.no

Abstract

This paper presents an approach on recognising individuals based on 3D acceleration data from walking, which are collected using MEMS. Unlike most other gait recognition methods, which are based on video source, our approach uses walking acceleration in three directions: vertical, backward-forward and sideways. Using gait samples from 21 individuals and applying two methods, histogram similarity and cycle length, the equal error rates of 5% and 9% are achieved, respectively.

1. Introduction

Mobile and portable electronic devices like mobile phones and PDAs, etc. are now an essential tool in people's everyday life. They are no longer considered as merely communication means, such devices can also be used in applications like m-banking [23] and m-government [13]. Consequently, they store and process valuable data, e.g. financial and private information. Thus, the devices can be targets of attackers not only because of their value per se but also because of their stored information. The present defense of these devices against unauthorised usage is usually based on PIN code and password, which are not always effective due to their security and memorability aspects [30]. To improve security in such devices, different biometric traits such as voice [17] and fingerprints [25, 26] have been proposed. However, unlike these biometrics, which are obtrusive and require user's attention, gait[1] biometric has an advantage of being non-invasive and obtained without a walker's attention. Studies from medicine [20] and psychology [10] present an evidence for considering human gait as having distinctive patterns from which individuals can be recognised. Despite the fact that gait biometric is relatively recent compared to, for example fingerprints, there are many studies devoted to gait recognition [8, 16, 12, 9, 22, 27, 11, 15, 19, 7, 4, 21]. All of these studies use machine vision techniques to extract gait patterns.

This paper presents a new method to recognise individuals by using their gait acceleration data. These accelecation data are obtained from a Micro-Electro-Mechanical System (MEMS), which is the integration of mechanical elements, sensors and electronics on a common framework [28, 14]. The approach uses acceleration data in three directions, namely vertical, backward-forward and sideways. In [1, 18] another approach of gait recognition that uses acceleration data is described. Our approach uses different positioning of the accelerometer device and tests a different type of gait signal. We attach the device to the leg, right above the ankle because it has more movement compared to the waist when a person walks. Additionally, the device we use for collecting data records acceleration signals at lower rate than used in [1, 18].

The rest of the paper is structured as follows: section 2 describes Gait Collection device and gait signal used, section 3 outlines applied methods, section 4 discusses experiment and its results, and section 5 concludes the paper.

2. Technology

2.1. The Gait Collection device

The device used to collect gait data for the analysis methods consists of an AVR Butterfly evaluation board from Atmel Corporation [3], which was equipped with two ADXL 202 dual axis accelerometers from Analog Devices [2]. The accelerometers were positioned perpendicularly to each other, making it possible to detect acceleration in three directions: vertical, backward-forward and sideways,

[1] Gait is a persons manner of walking.

as shown in Figure 1. The device was also equipped with an RS-232 interface for data transfer, 4.5 voltage external battery supply, and had a 4 Mbit dataflash for storage of acceleration data as shown in Figure 2. The Gait Collection device records acceleration at the rate of 16 samples per second.

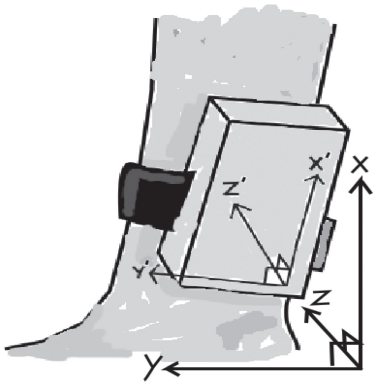


Figure 1. Acceleration directions when attached to leg.

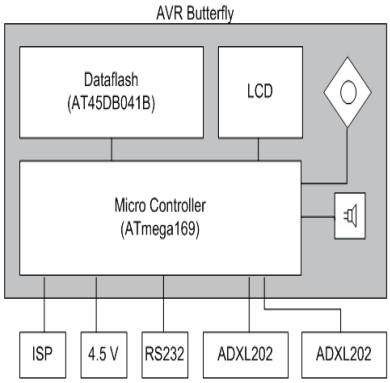


Figure 2. Diagram of Gait Collection device.

2.2. Acceleration signals

From the output of Gait Collection device acceleration signals in three directions: vertical X, backward-forward Y, and sideway Z are obtained. Instead of considering these raw acceleration signals separately, which might be sensitive to the device's attachment, we use a combined signal. Several combination of these signals are tested and the combined gait signal, which is constructed as follows, has shown best performance:

$$C_i = arcsin(\frac{Z_i}{\sqrt{X_i^2 + Y_i^2 + Z_i^2}}), i = 1, ..., k$$

where k is the number of recorded samples. The combined gait signal represents the alignment of the resultant gait signal (i.e. $\sqrt{X_i^2 + Y_i^2 + Z_i^2}$) to sideway axis. An example of acceleration signals in three directions and the combined gait signal is shown in Figure 3.

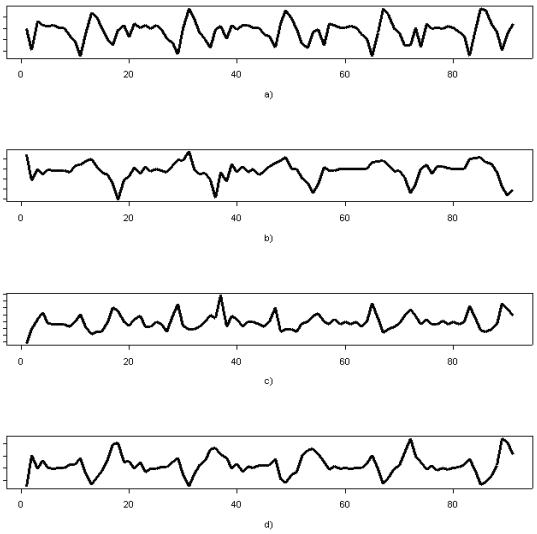


Figure 3. Fragments of gait acceleration signals: a) vertical X, b) backward-forward Y, c) sideways Z, and d) combined C.

3. Gait recognition methods

3.1. Cycle length method

The method is based on comparison of the gait cycle groups. In order to do this, the cycles are first detected and then the observation points inside the cycles are separated to the cycle groups. The cycle groups represent the gait cycles as populations of each general observation point sets. The idea of the comparison is to calculate the similarity scores between corresponding groups and create the similarity vec-

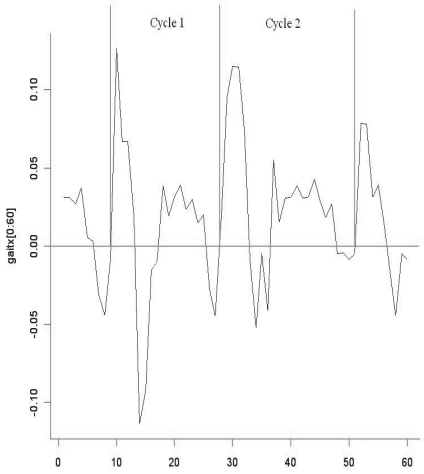


Figure 4. Cycles.

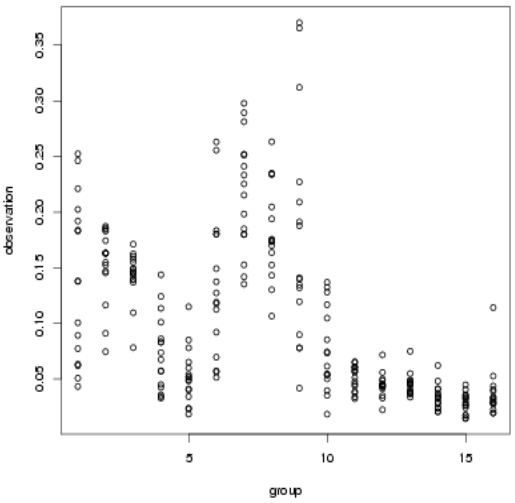


Figure 5. Cycle groups.

tor. The final comparison score is then determined based on the similarity vector.

Finding the cycles: The cycles are easiest to detect from the vertical acceleration signal. The data is first scaled by subtracting from each observation the mean value of the whole data and the zero values are observed, Figure 4.

The observations rarely hit to a zero, so the point where acceleration is zero is decided to be a negative value which is followed by a positive value. The first zero point is detected visually from the data. Other zero points are then detected automatically with help of the information of the cycle length. The cycle length can be estimated by using the autocorrelation function. After that the combined gait signal is divided to the cycles based on observations made from the vertical acceleration data.

Cycle groups: The number of groups depends on the cycle length. The longer the cycle is, the more observation points there are in one cycle. In this experiment the cycle length varies from 16 to 22 observations per cycle. In order to compare all individuals, only the first 16 observations from each cycle are used for comparison. This means that 16 groups are constructed. The first observation points (points where acceleration is zero) from each cycle are collected to the first group, the second observations from each cycle created the second group, etc. An example of the groups is presented in Figure 5.

Comparison of groups: The ideal situation is when the person walks similarly all the time, i.e., with constant speed and walking pattern. This would lead to equal gait cycles and, therefore, normally distributed gait groups because the values of the observation points at the same phase of the curve would be very close to each others. The variance of the cycle group would be very small and the mean value of the group could be used for comparison of the two different gait signals. Even though our experiment data is far away from the ideal situation, the mean of the groups are still tested for comparison. The similarity of the equal mean is calculated by the *t-test* of the statistical program R [24]. The variances of the population are considered to be unequal and that is taken account in the settings used in the *t-test*. After comparison between two datasets, the comparison vector contains 16 probability scores. The final score of the comparison is the total number of the probabilities, which score 27% or higher.

3.2. Histogram similarity

A n-bin histogram of the combined gait signal is computed. Then, histograms are normalised by the number of recorded samples. As the distance metric between two histograms we use the absolute distance,

$$dist(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

Here x_i is the probability of a data point falling into bin i of the enrollment's normalised histogram and y_i is the prob-

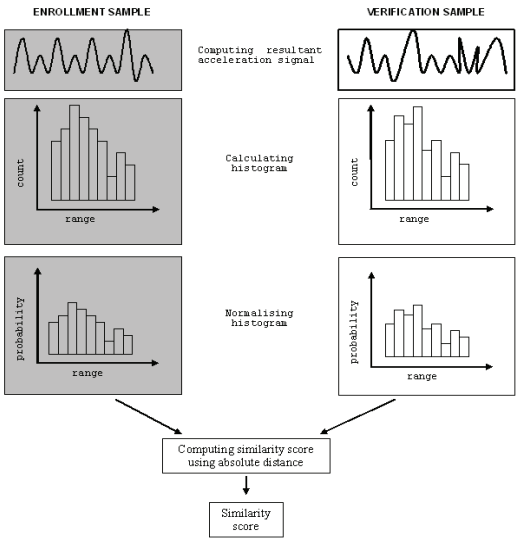


Figure 6. The process of applying histogram simiraity method.

Figure 7. Attachment of the Gait Collection device.

ability of a data point falling into bin i of the verification's normalised histogram.

The process of comparing two gait signals using histogram similarity is visualised in Figure 6.

4. Experiment and results

4.1. Experiment

Both analysis methods described were applied to the same sets of gait data. These data sets were collected using a population of 21 participants; 12 male and 9 female aging between 20 and 40 years old. The Gait Collection device was attached to the participants' right leg as shown in Figure 7. They ensured the device was firmly attached before their walking trial, so it did not shift much.

The participants walked in their normal walking speed a total distance of about 70 meters. The obtained gait data from these walking trials were divided into two parts to create two data sets for each person. The first set acted as an enrollment sample during the data analysis, while the second one acted as the person's verification sample. When the data analysis methods were tested, each of the enrollment samples was compared to each of the verification samples. This way, it was possible to simulate one genuine verification trial and 20 impostor trials for each participant, which in total generated 21 genuine and 420 impostor trials.

4.2. Results and discussion

Two different methods were tested for the comparison of the combined gait signals. The Equal Error Rates (EER) for the cycle length and the histogram methods are 9% and 5%, respectively. The DET-curves are shown in Figure 8.

Our results support the study made by Mäntyjärvi et al. [18]. Both studies were based on acceleration data of human gait even though there were differences between the studies. In the study of Mäntyjärvi et al. [18] the accelerations were measured from the waist of the test persons, while we collected acceleration data from the ankle of the participant. The placement of the device makes the detected acceleration curves different. Mäntyjärvi et al. [18] took account two acceleration signals while we used all three. The sample set varied also between studies. The correlation, FFT, and two variations of data distribution statistic methods of Mäntyjärvi et al. [18] achieved the EER of 7%, 10%, 18% and 19%, respectively.

Their correlation method and our cycle length method are both based on the gait cycles and, despite the differences in total analysis method and dataset, the EER figures are close to each other (7% vs. 9%). The bigger difference can be noted between their and our histogram methods (5% vs. 18%). The difference might be explained that we used the smaller data set and the fact that the data was collected at the same day.

When comparing MEMS based gait recognition methods with machine vision techniques, it can be noted that the EER figures are at the same level and comparable with each

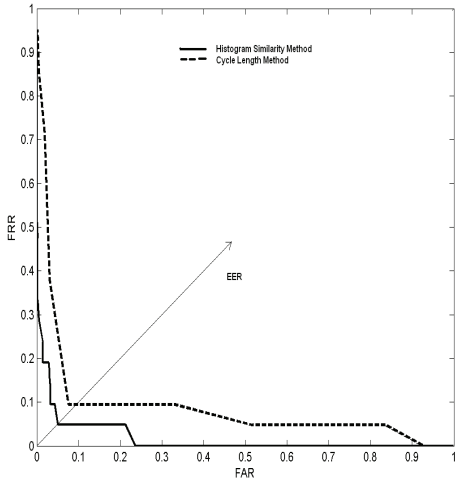


Figure 8. DET curve.

others. In studies [6, 5, 29], which all are based on video signals the EER are between 8-24%.

5. Conclusion

Our results confirm the possibility of recognising individuals based on their walking acceleration data. Although experimental results with a small number of subjects are promising, however, further studies are required. This is necessary to develop methods, which are more robust to changes such as footwear and surface. Current research on gait recognition indicates that it cannot be proposed for high security applications. However, when combined with other authentication mechanisms (e.g. PIN) gait biometric might enhance security and even improve the usability of the system. For example, in some situations it might be preferable to walk few steps instead of recalling rarely used passwords to activate a mobile device.

Even though MEMS based gait recognition lacks some difficulties of video based gait recognition, e.g. lighting conditions or background subtraction, it shares common challenges of gait recognition such as type of footwear, surface and injuries. The application area of MEMS based gait recognition differs from video based gait recognition. Applications of video based gait recognition systems are usually surveillance and forensic ones. Application area of MEMS based gait recognition, for instance, can be protection of mobile and portable electronic devices and access control.

We plan to perform experiments on a larger population of people and check the effect of the time factor for the performance of gait recognition. The current prototype of the device we used to collect acceleration data is not so convenient and it is an ongoing work on developing a more compact, portable, and user friendly version of it.

References

[1] H. Ailisto, M. Lindhold, J. Mäntyjärvi, E. Vildjiounaite, and S. Mäkelä. "Identifying people from gait pattern with accelerometers". In *Proceedings of SPIE Volume: 5779; Biometric Technology for Human Identification II*, pages 7–14, March 2005.

[2] Analog Devices. "ADXL202 data sheet". http://www.analog.com/UploadedFiles/Data_Sheets/53728567227477ADXL202E_a.pdf, Last visit: 20.01.2006.

[3] Atmel Corp. "AVR Butterfly evaluation kit: User guide". http://www.atmel.com /dyn/resources/prod_documents/doc4271.pdf, Last visit: 20.01.2006.

[4] C. BenAbdelkader, R. Cutler, and L. Davis. "Person identification using automatic height and stride estimation". In *16th International Conference on Pattern Recognition*, pages 377 – 380, 2002.

[5] C. BenAbdelkader, R. Cutler, and L. Davis. "Stride and cadence as a biometric in automatic person identification and verification". In *IEEE International Conference on Automatic Face and Gesture Recognition*. Microsoft Research, 2002.

[6] A. Bobick and A. Johnson. "Gait recognition using static activity-specific parameters". In *Proceedings of Computer Vision and Pattern Recognition Conference (CVPR 2001)*, Kauai, Hawaii, December 2001.

[7] J. Davis and S. Taylor. "Analysis and recognition of walking movements". In *International Conference on Pattern Recognition*, pages 315–318. Quebec City, Canada, August 2002.

[8] J. Hayfron-Acquah, M. Nixon, and J. Carter. "Automatic gait recognition by symmetry analysis". In *Audio- and Video-Based Biometric Person Authentication*, pages 272–277, 2001.

[9] J. Herrero-Jaraba, C. Orrite-Urunuela, D. Buldain, and A. Roy-Yarza. "Human recognition by gait analysis using neural networks". In *Proceedings of the International Conference on Artificial Neural Networks*, volume 2415, pages 346–369, 2002.

[10] G. Johannson. "Visual perception of biological motion and a model for its analysis". *Perception and Psychopsysics*, pages 201–211, 1973. 14.

[11] A. Johnson and A. Bobick. "A multi-view method for gait recognition using static body parameters". In *3rd International Conference on Audio- and Video-Based Biometric Person Authentication*, Halmstad, Sweden, June 2001.

[12] A. Kale, A. Sundaresan, A. Rajagopalan, N. Cuntoor, A. RoyChowdhury, V. Krueger, and R. Chellappa. "Identification of humans using gait". *IEEE Transactions on Image Processing*, 13(9):1163–1173, September 2004.

[13] Y. Kim, J. Yoon, S. Park, and J. Han. "Architecture for implementing the mobile government services in Korea". In *Conceptual Modeling for Advanced Application Domains: ER 2004 Workshops CoMoGIS, CoMWIM, ECDM, CoMoA, DGOV, and eCOMO*, Shanghai, China, November 2004.

[14] R. Lal, P. R. Apte, K. N. Bhat, G. Bose, S. Chandra, and D. K. Sharma. "T7: MEMS: Technology, design, CAD and applications". In *ASP-DAC*, pages 24–25, 2002.

[15] L. Lee and W. Grimson. "Gait analysis for recognition and classification". In *IEEE Conference on Face and Gesture Recognition*, pages 155–161, 2002.

[16] L. Lee and W. Grisom. "Gait appearance for recognition". In *International ECCV 2002 Workshop on Biometric Authentication*, pages 143–154, 2002.

[17] Y. Lee, C. Seo, J. Lee, and K. Lee. "Speaker verification system for PDA in mobile-commerce". In *Web Communication Technologies and Internet-Related Social Issues - HSI 2003, Second International Conference on Human Society@Internet*, Seoul, Korea, June 2003.

[18] J. Mäntyjärvi, M. Lindhold, E. Vildjiounaite, S. Mäkelä, and H. Ailisto. "Identifying users of portable devices from gait pattern with accelerometers". In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 973–976, 2005.

[19] S. Mowbray and M. Nixon. "Automatic gait recognition via Fourier descriptors of deformable objects". In *Proceedings of Audio Visual Biometric Person Authentication*, pages 566–573. Guildford, 2003.

[20] M. Murray. "Gait as a total pattern of movement". *American Journal of Physical Medicine*, pages 290–332, 1967. 46(1).

[21] M. Nixon, J. Carter, J. Nash, P. Huang, D. Cunado, and S. Stevenage. "Automatic gait recognition". In *IEEE Colloquium on Motion Analysis and Tracking*, pages 3/1–3/6, London, UK, 1999.

[22] S. Niyougi and E. Adelson. "Analyzing and recongizing walking features in XYT". In *IEEE Computer Society International Conference on Computer Vision and Pattern Recongition*, pages 469–474, June 1994.

[23] K. Pousttchi and M. Schurig. "Assessment of today's mobile banking applications from the view of customer requirements". In *37th Annual Hawaii International Conference on System Sciences (HICSS'04)*, 2004.

[24] R Development Core Team. "R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria". http://www.r-project.org/, Last visit: 17.11.2005. ISBN 3-900051-07-0.

[25] Q. Su, J. Tian, X. Chen, and X. Yang. "A fingerprint authentication mobile phone based on sweep sensor". In *Third International Conference on Advances in Pattern Recognition (ICAPR)*, Bath, UK, August 2005.

[26] Q. Su, J. Tian, X. Chen, and X. Yang. "A fingerprint authentication system based on mobile phone". In *5th International Conference on Audio- and Video-Based Biometric Person Authentication*, July 2005.

[27] R. Tanawongsuwan and A. Bobick. "Performance analysis of time-distance gait parameters under different speeds". In *4th International Conference on Audio- and Video-Based Biometric Person Authentication*, Guildford, UK, June 2002.

[28] J. A. Walraven. "Introduction to applications and industries for microelectromechanical systems (MEMS)". In *International Test Conference*, pages 674–680, 2003.

[29] L. Wang, T. Tan, W. Hu, and H. Ning. "Automatic gait recognition based on statistical shape analysis". *IEEE Transactions on Image Processing*, Vol. 12, Issue 9:1120–1131, Sept 2003.

[30] J. Yan, A. Blackwell, R. Anderson, and A. Grant. "Password memorability and security: empirical results". *Security and Privacy Magazine*, Vol. 2, Issue 5:25–31, Sept.-Oct. 2004.

Chapter 10
Biometric Gait Authentication Using Accelerometer Sensor

Davrondzhon Gafurov, Kirsi Helkala, and Torkjel Søndrol

In *Journal of Computers* Vol.1, Issue 7, 2006

Biometric Gait Authentication Using Accelerometer Sensor

Davrondzhon Gafurov, Kirsi Helkala, and Torkjel Søndrol
Norwegian Information Security Lab - NISlab
Department of Computer Science and Media Technology
Gjøvik University College
P.O. Box 191, 2802 Gjøvik, NORWAY
{davrondzhon.gafurov, kirsi.helkala}@hig.no, mail@torkjel.com

***Abstract*— This paper presents a biometric user authentication based on a person's gait. Unlike most previous gait recognition approaches, which are based on machine vision techniques, in our approach gait patterns are extracted from a physical device attached to the lower leg. From the output of the device accelerations in three directions: vertical, forward-backward, and sideways motion of the lower leg are obtained. A combination of these accelerations is used for authentication. Applying two different methods, histogram similarity and cycle length, equal error rates (EER) of 5% and 9% were achieved, respectively.**

***Index Terms*— security, biometric, gait recognition, sensor-based gait, unobtrusive authentication, leg acceleration**

I. Introduction

With advances in miniaturization techniques, performance of the mobile and portable devices is rapidly increasing. This enables to use such devices not only as communication tools but also in an applications like m-banking [1] or m-government [2]. This means that they can store and process valuable information such as financial or private data. This also increases the risk of being target of attacks. According to UK statistics in every three minutes a mobile phone is stolen [3]. The current protection mechanisms of these devices are usually based on PIN codes or passwords. Nowadays a "heavy" user has on average 21 passwords to remember [4]. Unfortunately, 81% of the users select common words as a passwords and 30% of users write their passwords down, which equally compromises security [4]. Recently, biometric modalities such as fingerprints [5], [6] have been proposed for mobile devices. However, both fingerprints and password entry are obtrusive and require explicit action from the user, which is not convenient in a frequent use. In order to improve security in mobile and portable devices, an unobtrusive mechanisms of authentication is desirable.

This paper presents a biometric authentication of individuals based on their gait. Gait is a person's manner of walking. Unlike most of the previous gait recognition approaches, in our approach gait patterns are extracted from a physical device attached to the lower leg. From the output of the device accelerations in three directions: vertical, forward-backward and sideways motion of the lower leg are obtained. A combination of these accelerations is used for authentication.

Most of previous works in gait recognition are based on machine vision techniques, i.e. they process video or sequence of images to extract gait patterns. We will refer to this type of gait recognition as vision-based. Recently, a new direction in biometric gait recognition has emerged [7], [8], [9]. This direction significantly differs from vision-based methods in terms of technology. Instead of the camera, a physical device attached to the body is used for collecting gait patterns. We call this direction a sensor-based gait recognition. The work presented in this paper belongs to the sensor-based gait recognition group. Usually accelerometers are used as a sensor [7], [8], [9]. An inherent advantage of vision-based gait system is to capture gait of the person from the distance when other types of biometrics (e.g. fingerprints) are not available. A primary advantage of the sensor-based gait biometric over other type of biometric is that it enables unobtrusive user authentication.

The remainder of the paper is structured as follow section 2 briefly introduces biometric system and basic terms used in the paper, and section 3 presents previous work both on vision-based and sensor-based gait recognition. Section 4 contains a description of the the device and acceleration signals, while section 5 describes two applied methods. Section 6 presents experiments and results, and section 7 contains discussion and outlines some possible application areas for the sensor-based gait authentication. Finally, section 8 concludes the paper with the direction for future works.

II. Biometric system

Biometric systems operate by acquiring biometric data from an individual, extracting feature set from the acquired data, and comparing this feature set against the enrolment set in a database [10]. An enrolment sample of the user is assumed to have been previously obtained and

This paper based on "Gait Recognition Using Acceleration from MEMS" by D. Gafurov, K. Helkala and T. Søndrol which appeared in the Proceedings of the 1st IEEE International Conference on Availability, Reliability and Security 2006, Vienna, Austria, April 2006.

stored in the database. A verification sample of the user is the one which needs to be compared with enrolment sample(s) stored in the database in order to verify (or identify) the identity of the user. There are essentially two types of submitting biometric data, a genuine attempt and impostor attempt. The genuine attempt is a self-verification attempt when a user's submitted sample is compared to his own enrolment sample in the database. The impostor attempt is a nonself-verification attempt when user's verification sample is compared against another user's enrolment sample. A similarity or matching score is an output value from a recognition algorithm that represents how similar two biometric samples are. Based on the similarity score, the biometric systems decides whether to accept or reject a user.

Biometric systems operate in two modes: identification or authentication (also called verification). In the authentication mode, the system validates a person's identity by comparing the captured biometric data with his own biometric enrolled in the database. In this mode, the system conducts an one to one comparison. In identification mode, the system recognizes an individual by searching the enrolment samples of all users in the database for a match. In this mode, the system conducts one to many comparisons. In other words, the aim of authentication is to answer for the question "*Am I who I claim I am?*", while identification looks for the answer to the question "*Who am I?*". The focus of this paper is primarily on authentication.

III. Related work

A. Vision-based gait recognition

First studies on the recognition of human gait were reported from medicine and psychology [11], [12]. Johansson [11] demonstrated the ability of humans to recognize human locomotion from other types of motions. In addition, it was shown that people can also recognize friends from moving light displays (MLD) [12]. Barclay et al. [13] showed that humans can also recognize the gender of the walker from MLD. Earlier studies on vision-based gait recognition showed promising results, usually with small sample sizes [14], [15], [16], [17], [18], [19]. For instance, with a database of 16 gait samples from 4 subjects and 42 gait samples from 6 subjects, Hayfron-Acquah et al. [14] achieved correct classification rates of 100% and 97%, respectively. Furthermore, recent studies with a larger sample size (more than 100 subjects in the database) confirm gait as having discriminating power from which individuals can be identified [20], [21], [22], [23]. For example, Lam and Lee [20] achieved recognition rate over 80% with the database of 115 subjects and 2,128 walking sequences. Traditionally, gait recognition methods are grouped into two classes: model-based and model-free. Model-based approaches focus on recovering a structural model of the motion [18], [15], [24]. For example, in [18] static parameters of the body, such as the height, the distance between head and pelvis, the maximum distance between pelvis and feet, and the distance between feet are used for recognition. Model-free techniques aim to extract statistical features from a subject's silhouette [25], [26], [27]. For instance, Kale et al. [25] use the widths of the silhouette as a basic image feature, and then extract other gait features from it. Unlike model-free techniques, model-based approaches are, in general, view and scale invariant [28]. In a multi-biometric system, gait can improve performance of the system when combined with other type of biometrics, for example with face recognition [29], [30], [31] or foot ground reaction force [32]. For instance, in [29] recognition rates with gait and face profile separately were 85.7% and 64.3%, respectively. However, when they were integrated, performance increased up to 100%.

B. Sensor-based gait recognition

First works on sensor-based gait authentication were reported by Ailisto et al. [7] and Mäntyjärvi et al. [8]. They utilized acceleration of the waist for authentication. In our recent work [9], we investigated acceleration characteristics in hip movement for authentication. Sensor placements in [7] and [9] are shown in Figure 1 and Figure 2, respectively. Accelerometer sensors used in [7], [8] measure acceleration at the rate of 256 samples per second, while the sensor used in [9] records acceleration at the rate of about 100 samples/sec. Both in [7], [8] and [9] accelerometers record accelerations in three orthogonal directions. However, in [7], [8] accelerations only in two directions, forward-backward and vertical were used, while in [9] resultant acceleration of all three orthogonal directions were analyzed. As a way for unobtrusive multimodal biometric authentication, a combination of a sensor-based gait and voice biometric is proposed. Such fusion enabled to increase the performance of the recognition system in a noisy environment [33]. In addition, accelerometer sensor was tested in three different places, hip pocket, chest pocket and in the hand while walking [33], as shown in Figure 3.

Figure 1. Placement of the accelerometer sensor in the waist (used with permission from [7]).

Figure 2. Pen-resembling accelerometer sensor attached to the hip [9].

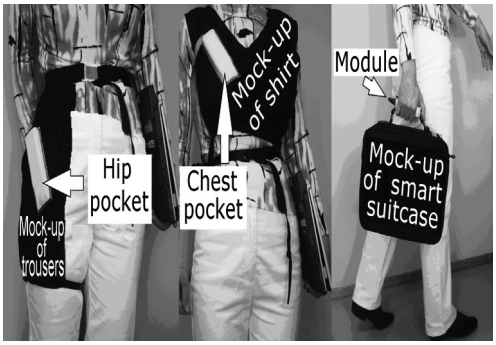


Figure 3. Accelerometer module in hip pocket, chest pocket and hand (used with permission from [33]).

IV. Gait Recognition Technology

A. The Motion Recording (MR) device

The Motion Recording (MR) device used to collect gait data for the data analysis consists of an AVR Butterfly evaluation board from Atmel Corporation [34] which has been equipped with two ADXL 202 dual axis accelerometers from Analog Devices [35]. The accelerometers have been positioned perpendicular to each other, making it possible to detect accelerations in three directions: vertical, backward-forward and sideways, as shown in Figure 4. An architecture of the MR device consists of a ATmega169 micro controller, an RS-232 interface for data transfer, a 4.5 Volt battery supply and a 4Mbit data flash for storage purposes as shown in Figure 5. The MR device is able to collect acceleration data between $\pm 2g$ ($g \approx 9.8 m/sec^2$) and has a sampling rate of 16 samples per second. The output from the accelerometers is a digital signal, whose duty cycles are proportional with the acceleration.

The Motion Recording device was encapsulated in a plastic box measuring 5.4 cm $\times$ 8.2 cm $\times$ 3 cm for protection, and straps were used to ensure a firm attachment to the leg.

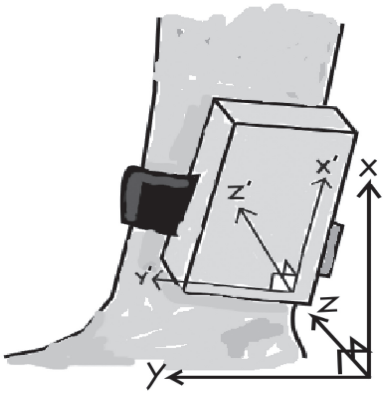


Figure 4. The MR device measures acceleration in three orthogonal directions: vertical, forward-backward and sideways.

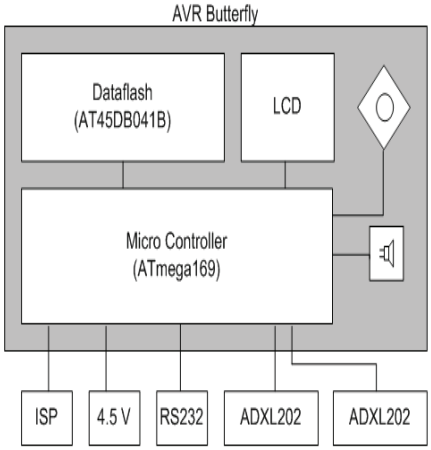


Figure 5. Architecture of Motion Recording device.

B. A combined acceleration signal

From the output of the Motion Recording device, acceleration signals in three orthogonal directions: X, Y, and Z are obtained. Instead of considering these raw acceleration signals separately, which might be sensitive to the device's attachment, we use a combined signal. Several combinations of these signals were tested and the combined gait signal, which is constructed as follows, has shown best performance:

$$R_i = arcsin(\frac{Z_i}{\sqrt{X_i^2 + Y_i^2 + Z_i^2}}), i = 1, ..., k$$

where X_i, Y_i, Z_i and R_i are vertical, forward-backward, sideway and combined acceleration at the observation number i; k is the number of recorded observations in the

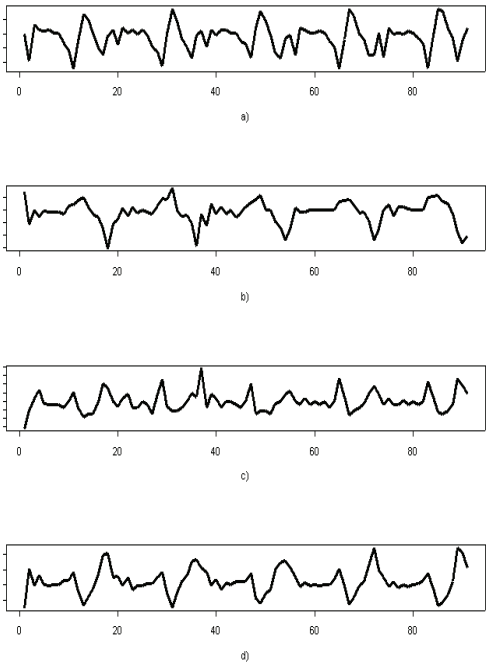


Figure 6. Fragments of gait acceleration signals: a) vertical X, b) backward-forward Y, c) sideways Z, and d) combined C.

signal. The combined gait signal represents the alignment of the resultant gait signal (i.e. $\sqrt{X_i^2 + Y_i^2 + Z_i^2}$) to the sideway axis (i.e. Z). An example of acceleration signals in three directions and the combined gait signal is shown in Figure 6.

V. Gait recognition algorithms

A. Histogram similarity

A n-bin histogram of the combined gait signal is computed. Then, histograms are normalized by the number of recorded observations. As the distance metric between two histograms we use the absolute distance,

$$dist(x, y) = \sum_{i=1}^{n} |x_i - y_i|.$$

Here x_i is the probability of a data point falling into bin i of the enrollment's normalized histogram x, and y_i is the probability of a data point falling into bin i of the verification's normalized histogram y. This distance value represents similarity score between two gait samples. Ideally, for genuine attempts the similarity scores should be smaller than for impostor attempts.

The steps involved in comparing two gait samples using the histogram similarity method are visualized in Figure 7.

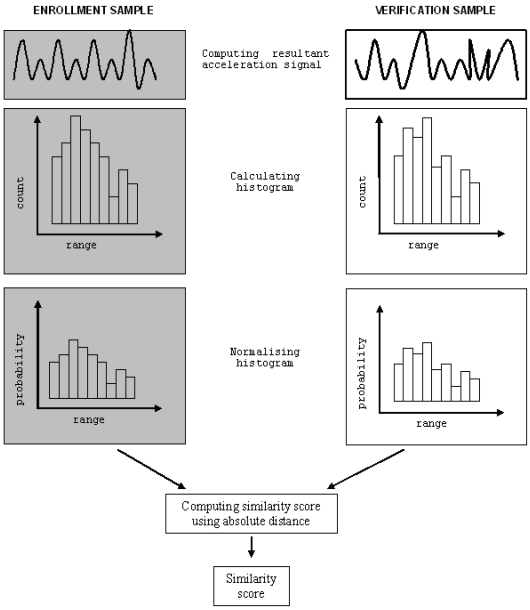


Figure 7. The process of applying the histogram similarity method.

B. Cycle length

The method is based on the comparison of gait cycle groups. The cycles are detected from the gait signal with help of the cycle length. The cycle groups are then created based on the observations point inside cycles. The cycle groups represent gait cycles as populations of general observation points. The idea of the comparison is to calculate similarity scores between corresponding groups and to create the similarity vector. The final comparison score is then determined based on the similarity vector.

Finding the cycles: The gait cycles are easiest to detect from the vertical acceleration signal. The data is first scaled by the formula

$$x_{i_f} = x_{i_o} - \bar{x},$$

where x_{i_f} is a scaled value, x_{i_o} is an observed value and $\bar{x}$ is the mean value of the data set. After scaling the starting points of each cycle are observed, as shown in Figure 8.

In theory the cycle starts at the point where the acceleration is zero. In practice the zero points are rarely found. Therefore the starting point is decided to be a negative value which is followed by a positive value. After the first starting point is detected the autocorrelation function is used to estimate the cycle length. By using the cycle length and observations made from the vertical acceleration data, the combined gait signal is divided to the cycles.

Cycle groups: The number of groups depends on the cycle length. The longer the cycle is, the more observation

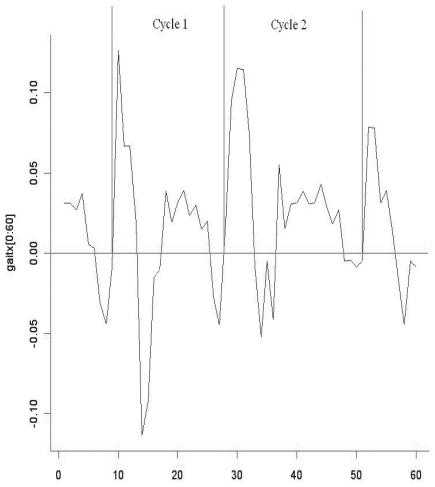


Figure 8. Gait cycles.

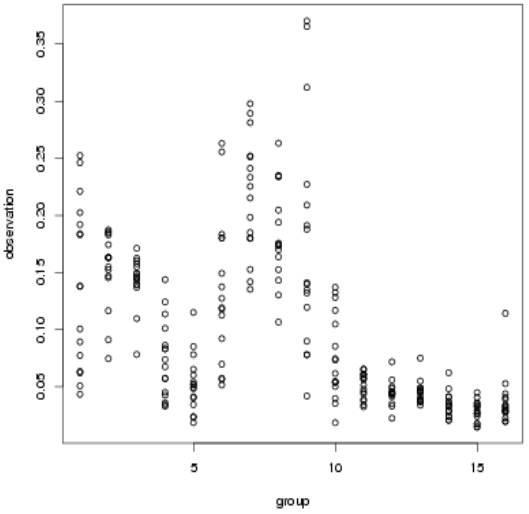


Figure 9. Cycle groups.

points there are in one cycle. In this experiment the cycle length varies from 16 to 22 observations per cycle. In order to compare all individuals only the first 16 observations from each cycle were used for comparison. This means that 16 groups were constructed for each individual. The first observation points (points where acceleration is zero) from each cycle were collected to the first group G_1, the second observations from each cycle created the second group G_2, and so on until G_{16}. An example of the groups is presented in Figure 9.

Comparison of groups: The ideal situation is that the person walks in a similar style all the time. His speed would be constant and his walking pattern would be the same. This would lead to equal gait cycles and therefore normally distributed gait groups because the values of the observation points at the same phase of the curve would be very close to each other. The variance of the cycle group would be very small and the mean value of the group could be used for comparison of the two different gait signals. Even though our experimental data represents a far from ideal data set, the mean of the groups were still tested for comparison.

The similarity of the equal means for each group G_i of persons A and B were calculated by two sample *t-tests*:

$$T_i = \frac{\bar{G}_{Ai} - \bar{G}_{Bi}}{\sqrt{\frac{s_{Ai}^2}{N_{Ai}} + \frac{s_{Bi}^2}{N_{Bi}}}}.$$

The variances s_{Ai} and s_{Bi} of the groups were considered to be unequal. The sample size N is the same in all groups. After comparison between two persons datasets, the statistic value vector T contains 16 probability scores. In order to compare the statistical vectors, the final score value S was calculated. The final score comparisons were made based on the probability 0.27,

$$S = \sum_{i=1}^{16} s_i,$$

where

$$\begin{cases} s_i = 1, \ if \ T_i \geq 0.27, \\ s_i = 0, \ otherwise. \end{cases}$$

The final similarity score S is then a value between 0 and 16.

VI. Experiment and Results

A. Experiment

Both analysis methods described were applied to the same sets of gait data. These data sets were collected using a population of 21 participants; 12 male and 9 female aged between 20 and 40 years. The MR device was attached to the participants' right leg as shown in Figure 10. Before the walking trials, the subjects ensured that the device was firmly attached, so it did not shift too much. They walked on a level and tiled surface in an indoor environment. After each walking trail collected acceleration data were transferred to the computer for analysis.

The participants walked at their normal walking speed a total distance of about 70 meters, during which they walked half of the distance in a straight line before turning around and walking back. The obtained gait data from these walking trials were manually divided into two parts by locating the section where the participants stopped and turned, to create two data sets for each person. In addition, the parts of the data sets which did not contain actual walking were removed. The first set acted as an enrolment

Figure 10. Motion Recording device attached to the lower leg.

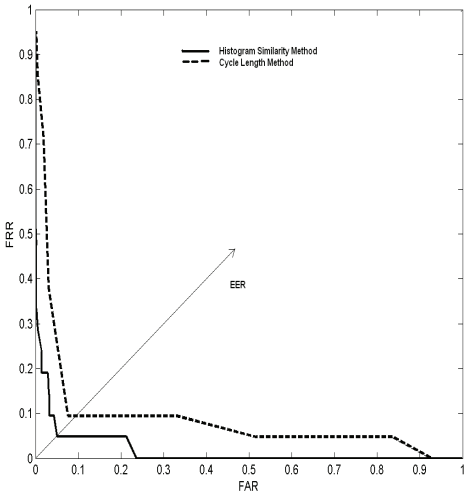


Figure 11. Performance of the methods in terms of the DET curves.

sample, while the second one was used as a verification sample. When the data analysis methods were tested, each of the enrolment samples was compared to each of the verification samples. This way, it was possible to simulate one genuine trial and 20 impostor trials for each subject, which in total generated 21 genuine and 420 impostor attempts.

B. Results

The performances of the methods in terms of Decision Error Trade-off curve (DET) are shown in Figure 11. The DET curve represents a plot of False Accept Rate (FAR) versus False Reject Rate (FRR), and characterizes performance of the biometric authentication system under different operating points (thresholds). For a given threshold, if the similarity score is less or equal to the threshold then the user is accepted, otherwise rejected. Error rates FAR and FRR are calculated as

$$FAR = \frac{N\_accepted\_impostors}{total\_N\_impostors},$$

and

$$FRR = \frac{N\_rejected\_genuines}{total\_N\_genuines},$$

respectively. In general, FAR relates to the security of the system, while FRR to the usability. An interesting point in the DET curve is the EER (equal error rate) where FAR=FRR. EER of the histogram similarity and cycle length are about 5% and 9%, respectively. For instance, an EER of 5% means that out of 21 genuine trials one is wrongfully rejected, and out of 420 impostor trials 21 are wrongfully accepted.

VII. Discussion and Application

A. Discussion

The mode of operation of the vision-based and sensor-based gait recognition systems is usually different. Sensor-based systems operate in authentication mode, while vision-based systems operate on identification mode. Therefore vision-based studies usually report performance of the recognition system in terms of classification rates [14], [15], [16], [18] or Cumulative Match Characteristic (CMC) curves [20], [36], [37]. Some of them alternatively report their results in terms of DET curves [36], [37], [38]. BenAbdelkader et al. [36] with a database of 17 subjects and using only stride and cadence as a biometric obtained EER of 11%. Wang et al. [37] using statistical shape analysis of the subjects' silhouettes achieved EER of 8%, 12% and 14% for three viewing angles (0^0, 90^0 and 45^0, respectively). Their database contained gait samples from 20 subjects. From the sensor-based gait recognition, Ailisto et al. [7] and Mäntyjärvi et al. [8] obtain EER of 7%, 10%, 18% and 19% for four different methods, namely signal correlation, frequency domain and two variations of data distribution statistics, respectively. Using time-normalized and averaged gait cycle method Gafurov et al. [9] obtained an EER of 16%. Performance of the multi-biometric authentication system proposed by Vildjiounaite et al. [33], where accelerometric gait and voice biometric was integrated, was between 2% and 12% EER depending on the level of background noise. Different characteristics of the current sensor-based gait recognition works are summarized in Table I. In the table, column two is the sensor's placement on the body, column three shows a sampling rate of the used accelerometer sensor, column four is the number of test subjects in the experiment, and the last column shows performance of the methods in terms of EER. Even though sampling rate of our accelerometer sensor is the lowest one, performances of our approaches are comparable with other works, and even better than some

TABLE I.
SUMMARY ON SENSOR-BASED GAIT RECOGNITION WORKS

	Placement	Rate, sam/sec.	N subjects	Performance, EER %
Ailisto et al. [7]	waist	256	36	6.4
Mäntyjärvi et al. [8]	waist	256	36	7, 10, 18, 19
Gafurov et al. [9]	hip	100	22	16
Vildjiounaite et al. [33] (gait+voice)	hip and chest pockets, hand	256	31	2-12
This paper	lower leg	16	21	5, 9

of them. In general, performances of the sensor-based and vision-based gait recognition systems are comparable on a small sample size. However, it should be noted that these comparisons are incommensurable as all the reported results are based on different data sets. The best way to compare different algorithms is to test them on the same sets of gait data.

Unlike, for example voice [39], [40] or handwritten signature [41], [42] biometrics, for which impersonation attacks have been studied, the security of gait biometrics has not received much attention. Only recently it has been hypothesized that minimal-effort impersonation attacks (mimicking someone's else walking style) on gait might not be successful. The hypotheses was based on the analysis of the passive and active impostor distributions [9]. Although this finding presents another advantage of the gait biometric, further analysis are required in this direction. For example, it is important to verify if impersonation attack can be improved by training of the hostile users; Are there such users whose walking style is relatively easy to mimic? Are there such attackers who can easily mimic other people? In Doddington et al. [43] terms, whether there are any "lambs" or "wolves" users in gait mimicking.

Though our aim was to investigate whether acceleration of the lower leg movement can be used for authenticating people, placement of the MR device has limitations from an application point of view. More appropriate placement of the device would be body segments used in [7], [9]. In the current version of the MR device the data transfer to the PC is done manually. Future versions of it should include a module for wireless communication, e.g. Bluetooth module. This may enable to conduct an on-line authentication of the users. In our approach, verification and enrolment samples were obtained in one session. In a realistic setting, they are obtained in separate sessions with at least few days time interval between sessions. In general, sensor-based gait recognition approach lacks difficulties of the vision-based system such as background subtraction, lighting conditions, viewing angles etc. Nevertheless, it shares common factors that can alter gait of the person like carrying load, surface, injury and so on. It should be noted that wearing Motion Recording device does not affect gait of the person significantly. Graves et al. [44] reported that lower extremity kinematics was insignificantly affected by the addition of loads 0.5 kg and 1 kg added to each foot. The Motion Recording device weighed about 350 gram.

B. Application

Generally, applications of the vision-based gait recognition focus on surveillance and forensics, whereas application of the sensor-based gait recognition system can be authentication and access control. For example, sensor-based gait biometrics have been proposed to improve security mechanisms in mobile devices [7], [8]. Another application area for sensor-based gait authentication system can be in the area of wearable computing. Wearable computers are computational devices that can be worn effortlessly, run continuously and be operated hands-free [45]. An iris-based user authentication is proposed for wearable computers [46]. Although performance of the iris-based user recognition is high, however, the iris-based systems require user cooperation. The issues of unobtrusiveness and user's attention are important in wearable computing environment [47]. Therefore gait can be a good candidate for authentication to wearable devices, provided that error rates can achieve satisfactory levels. Nevertheless, performance of the sensor-based gait system can be improved when it is combined with other types of authenticators (e.g. voice [33]), or using more than one sensor on different body parts.

VIII. CONCLUSION AND FUTURE WORK

A sensor-based gait recognition is a recent topic in the area of biometric gait recognition. In a sensor-based gait recognition approach, gait patterns are extracted using a sensor attached to the body. Despite technological differences between vision-based and sensor-based gait recognition system, their performances on small sample sizes are comparable. A primary advantage of sensor-based (or accelerometric) gait recognition over other type of biometric modalities is the ability to enable unobtrusive user authentication. In this paper we presented evidence towards sensor-based gait authentication by using acceleration of the lower leg. Although results with the small sample size are promising, however further work with larger sample size is necessary. It is also important to develop better algorithms to lower error rates. At the same time such algorithms should be robust against factors that can alter gait of the human, such as footwear, surface, carrying load, etc. Unlike vision-based systems, to our knowledge there is no established database for sensor-based gait recognition, which contains gait samples from at least 100 persons. Such database will allow to compare performances of different algorithms under common bases. All these topics will constitute the basis for our future work.

ACKNOWLEDGMENT

We would like to thank Prof. Einar Snekkenes at Gjøvik University College for providing the Motion Recording device, and Prof. Heikki Ailisto at Technical Research Center of Finland for allowing us to use Figures 1 and 3.

REFERENCES

[1] K. Pousttchi and M. Schurig, "Assessment of today's mobile banking applications from the view of customer requirements." in *37th Annual Hawaii International Conference on System Sciences (HICSS'04)*, 2004.

[2] Y. Kim, J. Yoon, S. Park, and J. Han, "Architecture for implementing the mobile government services in Korea," in *First International Workshop on Digital Government: Systems and Technologies (DGOV 2004)*, Shanghai, China, November 2004.

[3] "Huge surge in mobile phone thefts," http://news.bbc.co.uk/1/hi/uk/1748258.stm, Last visit: 04.09.2006.

[4] "2002 NTA Monitor password survey," http://www.out-law.com/page-3193, Last visit: 04.09.2006.

[5] Q. Su, J. Tian, X. Chen, and X. Yang, "A fingerprint authentication system based on mobile phone," in *5th International Conference on Audio- and Video-Based Biometric Person Authentication*, July 2005.

[6] X. Chen, J. Tian, Q. Su, X. Yang, and F. Y. Wang, "A secured mobile phone based on embedded fingerprint recognition systems," in *IEEE International Conference on Intelligence and Security Informatics*, May 2005.

[7] H. J. Ailisto, M. Lindholm, J. Mäntyjärvi, E. Vildjiounaite, and S.-M. Mäkelä, "Identifying people from gait pattern with accelerometers," in *Proceedings of SPIE Volume: 5779; Biometric Technology for Human Identification II*, March 2005, pp. 7–14.

[8] J. Mäntyjärvi, M. Lindholm, E. Vildjiounaite, S.-M. Mäkelä, and H. J. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2005.

[9] D. Gafurov, E. Snekkenes, and T. E. Buvarp, "Robustness of biometric gait authentication against impersonation attack," in *First International Workshop on Information Security (IS'06), OnTheMove Federated Conferences (OTM'06)*, Montpellier, France, Oct 30 - Nov 1, 2006, Springer LNCS, to appear.

[10] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics*, vol. 14, pp. 4–20, January 2004.

[11] G. Johansson, "Visual motion perception," in *Scientific Am.*, 1976, pp. 75–88, vol.32.

[12] J. Cutting and L. Kozlowski, "Recognizing friends by their walk: gait perception without familiarity cues," in *Bulletin of the Psychonomic Society 9*, 1977, pp. 353–356.

[13] C. Barclay, J. Cutting, and L. Kozlowski, "Temporal and spatial factors in gait perception that influence gender recognition," *Perception and Psychophysics 23*, pp. 145–152, 1978.

[14] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Automatic gait recognition by symmetry analysis." in *Audio- and Video-Based Biometric Person Authentication*, 2001, pp. 272–277.

[15] D. Cunado, M. Nixon, and J. Carter, "Automatic extraction and description of human gait models for recognition purposes," in *Computer Vision and Image Understanding*, 2003, pp. 1 – 41.

[16] L. Wang, W. Hu, and T. Tan, "A new attempt to gait-based human identification." in *International Conference on Pattern Recognition*, 2002, pp. 115–118.

[17] C. BenAbdelkader, R. Cutler, H. Nanda, and L. Davis, "Eigengait: Motion-based recognition of people using image self-similarity." in *Thrid International Conference on Audio- and Video-Based Biometric Person Authentication*, 2001.

[18] A. Y. Johnson and A. F. Bobick, "A multi-view method for gait recognition using static body parameters," in *Third International Conference on Audio- and Video-Based Biometric Person Authentication*, June 2001, pp. 301–311.

[19] J. Shutler and M. Nixon, "Zernike velocity moments for description and recognition of moving shapes." in *British Machine Vision Conference*, 2001, pp. 11/1 – 11/4, session 8: Modelling Behaviour.

[20] T. H. W. Lam and R. S. T. Lee, "A new representation for human gait recognition: Motion silhouettes image (MSI)," in *International Conference on Biometrics*, 2006, pp. 612–618.

[21] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanID gait challenge problem: Data sets, performance, and analysis." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–177, 2005.

[22] Y. Wang, S. Yu, Y. Wang, and T. Tan, "Gait recognition based on fusion of multi-view gait sequences." in *International Conference on Biometrics*, 2006, pp. 605–611.

[23] M. S. Nixon, T. N. Tan, and R. Chapella, *Human Identification Based on Gait*, D. D. Zhang and A. K. Jain, Eds. Springer, 2006.

[24] D. Wagg and M. Nixon, "On automated model-based extraction and analysis of gait." in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 11–16.

[25] A. Kale, N. P. Cuntoor, B. Yegnanarayana, A. N. Rajagopalan, and R. Chellappa, "Gait analysis for human identification." in *4th International Conference on Audio- and Video-Based Biometric Person Authentication*, June 2003, pp. 706–714.

[26] P. Huang, C. Harris, and M. Nixon, "Human gait recognition in canonical space using temporal templates." in *Vision Image and Signal Processing*, 1999, pp. 93–100.

[27] H. Murase and R. Sakai, "Moving object recognition in eigenspace representation: gait analysis and lip reading," in *Pattern Recognition Letters*, 1996, pp. 155–162.

[28] N. Boulgouris, D. Hatzinakos, and K. Plataniotis, "Gait recognition: A challenging signal processing technology for biometric identification," in *IEEE Signal Processing Magazine*, November 2005, pp. 78–90.

[29] X. Zhou, B. Bhanu, and J. Han, "Human recognition at a distance in video by integrating face profile and gait." in *5th International Conference on Audio- and Video-Based Biometric Person Authentication*, July 2005, pp. 533–543.

[30] X. Zhou and B. Bhanu, "Integrating face and gait for human recognition," in *Conference on Computer Vision and Pattern Recognition Workshop*, June 2006.

[31] A. Kale, A. Roychowdhury, and R. Chellappa, "Fusion of gait and face for human identification," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04).*, May 2004.

[32] P. Cattin, D. Zlatnik, and R. Borer, "Sensor fusion for a biometric system using gait," in *International Conference on Multisensor Fusion and Integration for Intelligent Systems*, August 2001, pp. 233 – 238.

[33] E. Vildjiounaite, S.-M. Mäkelä, M. Lindholm, R. Riihimäki, V. Kyllönen, J. Mäntyjärvi, and H. Ailisto, "Unobtrusive multimodal biometrics for ensuring privacy and information security with personal devices," in *Pervasive*, May 2006, pp. 187–201, Springer LNCS.

[34] "Atmel corp." http://www.atmel.com, Last visit: 29.08.2006.

[35] "Analog devices," http://www.analog.com, Last visit: 29.08.2006.

[36] C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification," in *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, May 2002, pp. 357–362.

[37] L. Wang, T. Tan, W. Hu, and H. Ning;, "Automatic gait recognition based on statistical shape analysis," in *IEEE Transactions on Image Processing*, Sept. 2003, pp. 1120 – 1131, volume 12, Issue 9.

[38] A. Bobick and A. Johnson, "Gait recognition using static, activity-specific parameters," in *Proceedings of the 2001 IEEE Computer Computer Vision and Pattern Recognition*, 2001, pp. I–423 – I–430, vol.1.

[39] J. Lindberg and M. Blomberg, "Vulnerability in speaker verification - a study of technical impostor techniques," in *Eurospeech*, 1999, pp. 1211–1214.

[40] Y. W. Lau, M. Wagner, and D. Tran, "Vulnerability of speaker verification to voice mimicking," in *International Symposium on Intelligent Multimedia, Video and Speech Processing*, October 2004, pp. 145 – 148.

[41] J. Guo, D. Doermann, and A. Rosenfield, "Off-line skilled forgery detection using stroke and sub-stroke properties," in *15th International Conference on Pattern Recognition*, September 2000, pp. 355 – 358.

[42] S.-H. Cha and C. Tappert, "Automatic detection of handwriting forgery," in *Eighth International Workshop on Frontiers in Handwriting Recognition*, August 2002, pp. 264 – 267.

[43] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, goats, lambs and wolves a statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation," in *5th International Conference on Spoken Language Processing*, 1998.

[44] J. Graves, A. Martin, L. Miltenberger, and M. Pollock, "Physiological responses to walking with hand weights, wrist weights, and ankle weights." in *Journal of Orthopaedics and Traumatology*, 1988, pp. 265–271, med Sci Sports Exerc. June:20(3).

[45] P. Huang, "Promoting Wearable Computing: A Survey and Future Agenda," in *In Proc. of International Conference on Information Society in The 21st Century: Emerging Technologies and New Challenges*, November 2000.

[46] J. J. Lee, S. Noh, K. R. Park, and J. Kim, "Iris recognition in wearable computer," in *First International Conference on Biometric Authentication*, July 2004, pp. 475–483.

[47] T. Starner, "Attention, memory, and wearable interfaces," *Pervasive Computing, IEEE*, vol. 1, Issue 4, pp. 88–91, Oct.-Dec. 2002.

Davrondzhon Gafurov received his MSc degree in computer engineering from Technological University of Tajikistan (TUT), Khujand, Tajikistan, in 2000. During 2000-2004 he worked as an engineer-programmer at the Computer Center of Technological University of Tajikistan (CCTUT), and at the same time he was a lecturer at the department of Programming and Information Technology at TUT. He is currently a Ph.D. candidate at the department of Computer Science and Media Technology at Gjøvik University College, Norway. His current research interest is in the area of information security, particularly biometrics systems and robustness of biometrics against attacks.

Kirsi Helkala received her MSc in mathematics from University of Joensuu, Finland, in 2001. After graduating she has worked as an assistant at the departement of Mathematics at University of Joensuu and then as a mathematics, physics and chemistry teacher in Kesämäenrinteen Koulu in Lappeenranta, Finland. She is currently a PhD student at Gjøvik University College, Norway. Her work interest is personnel authentication, especially in a health service context.

Torkjel Søndrol received his MSc degree in information security from the department of Computer Science and Media Technology at Gjøvik University College, Norway in 2005. He is currently engaged in a project trying to commercialise results from his MSc degree. His current research interests are in information security related to programming as well as authentication.

Chapter 11

Face Recognition Using Separate Layers of the RGB Image

Patrick Bours and Kirsi Helkala

In *Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Special Session on Biometrics - From Sensors to Standardization*, August 15-17, 2008, Harbin, China, pp. 1035-1042, IEEE Press, 2008

Face recognition using separate layers of the RGB image

Patrick Bours and Kirsi Helkala
Gjøvik University College
Norwegian Information Security Laboratory
patrick.bours@hig.no and kirsi.helkala@hig.no

Abstract

In many cases face recognition of still images is performed with greyscale images. These images are actually converted from a color image to greyscale before the analysis takes place. A consequence of such a conversion is obviously loss of information, which could influence the performance of the face recognition system. It would be interesting to see if using one of the three color layers of the RGB image could give better recognition performance compared to the greyscale converted image. We conducted two experiments and the results indeed support this idea. We found that the red layer of the RGB image gives the best recognition performance, especially in the cases where an extra light source is used to light up (part of) the face of the participants in the experiments. In the case that the participants were facing the camera we saw the Equal Error Rate drop from 3.3% for the greyscale images to 1.8% for the red layer of the RGB images in our initial experiment.

1. Introduction

Face recognition seems to be done effortless by human beings when recognizing friends, family, and other people. We have no problems recognizing people under many different circumstances like different illumination conditions, different facial expressions, partially blocked faces etcetera. For a computer such a task is much more difficult.

Face recognition has been studied for over 35 years and it has many applications. These applications can be separated roughly into two different classes: still-images and videos. An example of the use of still images is the verification of a driver's license and an example of video is room surveillance with comparison of detected individuals with a watch list. According to [25] face recognition is used in the following areas: entertainment, smart cards, authentication, information security, law enforcement and surveillance. It is a non-intrusive authentication method, can be used without co-operation of a person and it is highly accepted among users [8]. Even though face recognition is widely studied and used, there still remain several challenges. It is common sense that illumination, poses and expressions cause the most severe problems to receive accurate face recognition [25]. Our study is addressed toward one of these: illumination. In both our initial and our large scale experiment we used an extra light source to see the differences in results due to extra illumination. In our initial experiment we used 3 different poses (both frontal and facing left and right), but the resulting 3 sets of images will be treated separately. In the large scale experiment we only used the frontal pose but now with more volunteers and more images per volunteer.

A technique that is often applied in machine face recognition is so-called eigenfaces [19]. In such a case the images are regarded as vectors in a high dimensional space and the basis of this space is transformed in such a way that the new basis vectors are "face-like" images or so-called eigenfaces. The transformation is based upon a set of input images and is called Principle Component Analysis (PCA). This is explained in more detail in Section 4.2.

Most of the papers on face recognition with eigenfaces use greyscale pictures. In such cases the cameras capture color images that are transformed into grey images. We will show however that it might be better to not transform the images to greyscale but use the separate color layers of the original image. This conclusion is based on two experiments that leaves much room for future research and, as a consequence, the conclusions from this paper should be tested in different settings.

2 State of the Art

It is next to impossible to recognize an individual based on a picture from his face profile, when a frontal picture was generated as a reference [25]. Any pose variation between a frontal and the full profile pose will cause an increasing challenge to the face recognition algorithm. Gabor wavelet transform together with PCA are in use [4] and it was shown that continuous pose changing forms a smooth curve in the pose eigenspace. The first principal compo-

nent divides all poses into two symmetric parts: from profile to frontal and from frontal to profile. The second one differentiates between poses profile and frontal views and the third one contains also information about illumination. Li and Su combined PCA with support vector machine to estimate pose angles [7]. The recognition accuracy achieved with combined PCA and Support Vector Machine method, CPCA-SVM, and 1000 samples was 97%. This accuracy dropped down to 96% when only 50 samples were used.

By illumination we mean the lighting conditions on the face. If a bright light is directed straight to the face, the face can almost be non-shaded. If a single point light source is used and the light ray is coming from a 90 degree angle, the shadows are long and dark and one half of the face is almost fully in the shadow. Several approaches have been suggested to handle this. Shan et. al. [18] have investigated several illumination normalization methods. The goal of these methods is to make the whole face visible, so that the shadowed part of the face is also recognizable. They have used 64 frontal images, captured under different illumination conditions for ten persons from Yale database B [22]. The images were further divided into 5 disjoint subsets based on the illumination conditions. The evaluated methods were Histogram Equalization (HE) globally over the images, Region-based Histogram Equalization (RHE), Gamma Intensity Correction (GIC) globally, Region-based GIC (RGIC), Quotient Illumination Relighting (QIR) and then combinations of the previous ones: GIC+RHE, RGIC+RHE, RHE+RGIC and HE+RGIC. The best recognition accuracies, achieved when light source was up to 25° from camera axis, were 94.4% to 100%, but when the angle increases the accuracy radically dropped. Xie et. al. [21] on the other hand take also the shadows into account. They estimate illumination parameters, which are then used to relight an image.

Quite often greyscale images are used in face recognition [4, 6, 7, 18, 21, 25]. Images can be color images, but while processing the images for the actual recognition, color images are transformed to greyscale images. In image processing, colors are commonly used for face detection. Examples of this are Byrd et. al. [3], who used color-ratios for 2-D face detection . Yu et. al. in [24] presented a full dynamic face recognition model that uses color information to locate the position of the human face and its facial features. However, it was noted that brains use color information for face recognition. Yip and Sinha [23] state "*Color cues do play an important role in face recognition. Their contribution becomes evident when shape cues are degraded. [With blurry or faraway images], recognition performance with color images is significantly better than with greyscale images*". If the color is noted to contain important information for the human brain then it can help in computer vision-based face recognition. Rajapakse et.al. [17] used color images for actual recognition not just detecting the face features. They used Non Negative Matrix Factorization (NMF) to recognize color face images from the AR database [10]. The frontal color images in this database have large variations in facial expressions and illumination. They showed improved accuracy of color image recognition over grey level image recognition. Recognition accuracies for images with extra illumination, when neutral images were used as training set, were 86.5% for color and 85.0% for greyscale images. The accuracies were 94.5% and 93.5% for color and greyscale images, respectively, when an illumination set was used as a training set.

There are many different face recognition algorithms available and many of them have been evaluated by FERET (Face Recognition Technology [9, 13, 14]), FRVT (Facial Recognition Vendor Test) in 2000 [2], 2002 [12] and 2006 [15], and FRGC (Face Recognition Grand Challange [11]) evaluations. According to [25], the three top methods are: The Elastic Bunch Graph Matching (EBGM) system, the subspace LDA (Linear Discriminant Analysis) system, and the probabilistic eigenface system. PCA can be used alone or together with other method like done in [4, 7].

3 Experiment Setup

3.1 Initial Experiment

As part of the authentication course taught at Gjøvik University College, the students had to perform a group project on evaluating a biometric authentication method. To guide the students we performed a small example experiment, investigating the performance of face recognition based on separate color layers of an RGB image. For our experiment we used 30 volunteers, 20 male and 10 female, all students and employees of Gjøvik University College. All of the volunteers are Caucasian, except one who has a more Arabic/Asian appearance. Each volunteer was photographed 24 times, 8 times facing the camera, 8 times facing left and 8 times facing right. In every set of 8 pictures, the first 5 were without any extra light, while the last three had an extra light source, coming from the left, the front, respectively the right side of the camera. All other conditions were kept as constant as possible. The volunteers were asked to look as neutral as possible, have their mouth closed and take of their glasses if appropriate. All the pictures were taken in the same location, an inside room, against a light background. In Figure 1 we can see 12 (out of the 24) images of one volunteer. In this figure, the first column has no extra light source. Columns 2 to 4 show extra light from the left, the front, and the right side.

The normal light came from fluorescent strip lights at the ceiling of the room and the volunteers were placed exactly in the middle of two such light sources to provide an equally

Figure 1. Set of images of a volunteer

spread of the light. Measurements of this light revealed that it had a temperature of 2890K. The spectral graph of this light source showed three peaks, for the red, green and blue light, where the red peak was the highest. Furthermore, the height of the green peak was almost 60% of the height of the red peak and the height of the blue peak was approximately 25% of the red one. The natural sun light came from the window, approximately 5.20 meters to the right of the volunteers. The windows were almost completely blocked to minimize influences of different outside lighting conditions on the images. The extra light source we used was a 300W film lamp that had a temperature of 3174K and there were no special peaks in the spectral graph of this light source.

All volunteers were standing in the exact same position while being photographed and were given oral guidance on what to do. The distance to the camera was 2.3 meters, which was also the distance to the extra light source when it was used. The extra light source was placed at approximately 1.75 meters high, so around the same height as the face of the volunteers.

3.2 Large Scale Experiment

We also performed a larger experiment, only focusing on the volunteers when they faced the camera. In this experiment each volunteer was photographed 28 times: 7 times without extra light, 7 times with extra light from the left, 7 times with extra light from the front and finally 7 times with extra light from the right. We used 102 volunteers for this experiment, being 32 female and 70 male, all students or staff from Gjøvik University College, and a mixture of Caucasian and non-Caucasian. In this experiment the volunteers were again asked to remove their glasses, but now they were asked to talk to us during the session so that there was a larger variety in the facial expressions. The light sources were the same as in the initial experiment.

4 Data Analysis for the Initial Experiment

4.1 Preprocessing Facial Images

All pictures were taken with a Nikon D200 camera, storing the images in the Nikon's proprietary NEF (Nikon Electronic Format) format. These images were cropped manually to only contain the face of the volunteer, and finally the resulting images were resampled to contain exactly 100x100 pixels each. Before analysis the images were transformed to Portable PixMap (PPM) format, as this format is required by the R statistical analysis tool [16] for further analysis.

Reading a PPM image in R results in a structure that can be used to extract the different color layers. Each of the three layers is (in our case) a 100x100 dimensional array, which in fact will be treated as vector of length 10.000. So in fact, each image is regarded as a set of 3 vectors, each of length $n = 10.000$. For the analysis we use only a single vector of length n, being either one of the three original vectors (in the case where we considered the separate color layers of the image), or a combination of the 3 original vectors into one new vector.

A color image can be converted to a greyscale image by taking a linear combination of the three layers where the total weights of the factors sum up to 1. We have chosen to consider two of these conversions. The first conversion is having all the three weights equal to 1/3, meaning all colors have the same contribution to the grey image. In other words, if a pixel in the red layer has a value r, in the green layer has value g and in the blue layer has value b, then that particular pixel in the grey image will have value $y = (r + g + b)/3$. The second conversion method uses different weights for the three layers. This conversion is based on the Y-layer of NTSC YIQ color space [20], which is a greyscale version of an image. In [20] we find that the conversion equals $y = 0.30 \cdot r + 0.59 \cdot g + 0.11 \cdot b$. This conversion is amongst others used in the image processing program IrfanView [5]. In our analysis we used both conversion methods. We will denote the two conversions by "Grey1" (using the equal weights) and "Grey2" (using the YIQ conversion).

From here on when we talk about an image we will indicate either one of the three color layers or one of the two greyscale versions. In general we will make no distinction between an image with 100x100 pixels and a vector of length $n = 10.000$.

4.2 PCA and eigenfaces

When applying Principle Component Analysis (PCA) the basis vectors are transformed, based on an input data set. The transform of the basis is in such a way that the

new basis vectors, called Principle Components, better describe the structure of the underlying input data set. This technique was first used by Turk and Pentland in the area of face recognition [19]. In case of face recognition the Principle Components are called eigenfaces because they resembled faces.

Let $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m\}$ be the input dataset on which the PCA will be based, where each of the $\mathbf{x}_i$ is an n-dimensional vector. The first step in PCA is to compute the mean value $\mathbf{x}_{av}$ of the input set and subtract this mean from each of the $\mathbf{x}_i$. The new set is $\{\mathbf{x}'_1, \mathbf{x}'_2, ..., \mathbf{x}'_m\}$, where $\mathbf{x}'_i = \mathbf{x}_i - \mathbf{x}_{av}$. The vectors $\mathbf{x}'_i$ are now used as columns in an $n \times m$ matrix X and the covariance matrix $C = \frac{1}{n^2-1} \cdot X^T \times X$ is computed from X. Now C is a symmetric $n \times n$ matrix from which we compute the eigenvectors $\mathbf{ev}_i$ and the corresponding eigenvalues λ_i. The relation between the matrix C and the eigenvectors and eigenvalues is such that $C \times \mathbf{ev}_i = \lambda_i \cdot \mathbf{ev}_i$. The new basis will be formed by the eigenvectors $\mathbf{ev}_i$.

In case of face recognition, the input vectors are derived from a set of input images of faces. An image with $n_1 \times n_2$ pixels can obviously also be regarded as a vector of length $n := n_1 \cdot n_2$. Taking m input images and regarding them as vectors gives the input set for PCA. In this case the resulting eigenvectors are called eigenfaces. The reason for this is that these eigenvectors, regarded again as an $n_1 \times n_2$ pixel image, look like faces.

When we analyze the data, we will consider one pose at the time. Each of the 30 volunteers had 8 images for each of the three poses and one of these images will be used to build the PCA. In our analysis the eigenfaces are based upon the first image of each person, i.e. without any extra light source. So our dataset for PCA consisted of 30 images, each 100×100 pixels large. This means that our matrix X has size $n \times m = 10.000 \times 30$ and the covariance matrix C has size 10.000×10.000, which is too large to use to compute the eigenvalues. To overcome this problem we computed the PCA on the matrix X^T. The covariance matrix in this case equals $C' = \frac{1}{m^2-1} \cdot X \times X^T$, which is in our case a 30×30 matrix. For the matrix C' it is easy to compute the eigenvectors $\mathbf{ev}'_i$ and eigenvalues λ'_i, for $i = 1, 2, ..., 30$. The first 30 eigenvectors of the matrix C can now be found from the eigenvectors of the matrix C'. Note that by definition we have

$$C' \times \mathbf{ev}'_i = \frac{1}{m^2-1} \cdot X \times X^T \times \mathbf{ev}'_i = \lambda'_i \cdot \mathbf{ev}'_i$$

so we find

$$\frac{1}{n^2-1} \cdot X^T \times X \times (X^T \times \mathbf{ev}'_i) = \frac{m^2-1}{n^2-1} \cdot \lambda'_i \cdot (X^T \times \mathbf{ev}'_i).$$

Because $C = \frac{1}{n^2-1} \cdot X^T \times X$, we see that $X^T \times \mathbf{ev}'_i$ are scalar multiples of the eigenvectors of the larger system that we are looking for. So we are able to compute the first $m = 30$ eigenfaces and work with these. Obviously if we have an eigenvector $\mathbf{ev}$ of C then all scalar multiples of $\mathbf{ev}$ are also eigenvectors. We will rescale the m eigenvectors $\mathbf{ev}_i$ so that they all have norm 1, i.e. the innerproduct of $\mathbf{ev}_i$ with itself equals 1.

4.3 Reference Generation

In order to generate a biometric reference for each person we compute PCA-based templates from a single picture of the person, where the lighting is normal, i.e. no extra light source. The image is decomposed into the 3 vectors of length 10.000 related to the three color layers. Either one of these vectors, or a greyscale combination of them, is used to build the feature vector (template). The template consists of 30 values.

Let $\mathbf{y} = (y_1, y_2, ..., y_n)$ be the vector representation of the image needed to build the template. First the average face $\mathbf{x}_{av}$ is subtracted and next the resulting vector $\mathbf{y}'$ is expressed in the m eigenfaces. This is done by taking the innerproduct between the eigenvector $\mathbf{ev}_i$ and the vector $\mathbf{y}'$. The resulting numerical value should be actually be divided by the norm of the eigenvector, but the eigenvectors already have norm 1, so this is not needed. We now have transformed the image of user i into an m-valued template $T_i = (t^i_1, t^i_2, ..., t^i_{30})$.

4.4 Analysis Setup

The analysis is performed as follows. From each volunteer we had 8 pictures facing one direction (front, left and right). We will only consider one of these directions at a time. Out of the 8 images 5 were with no extra light source. Image number 6 had extra light from the left, number 7 from the front and number 8 from the right. When doing the analysis, we do not only restrict ourselves to pictures facing one direction, but also to either one color layer or a greyscale version of the images. This means that we use similar information when building the PCA, creating the templates and transforming the images for analysis.

The eigenvectors were created with image number 1 of each volunteer and the template was build from image number 2 of each person. In the analysis we varied on the one hand the input images and on the other hand the information extracted from each of these images. We used 6 sets of input images, being images 3-5, image 6, image 7, image 8, images 6-8 and images 3-8. The other variation was in the extracted information which was either the red layer, the green layer, the blue layer, a greyscale version with each layer represented for 1/3 (grey1) or a greyscale version according to the YIQ conversion (grey2).

All of the images that are used in the analysis are transformed into m-valued inputs in the same way as is described in Section 4.3 for the templates. In this case the template and the input feature vector are in a comparable format. For all of the $3 \cdot 6$ different combinations of input images and extracted information we computed the Euclidean distance between the input and the templates. The Euclidean distance metric was used because of its simplicity and the fact that initial tests showed that this distance metric gave a good performance. The distance information is used to find the False Acceptance Rates (FAR) and False Rejection Rates (FRR) for various thresholds. From this information the Equal Error Rate (EER) was determined. These EER are reported in Section 5.

5 Results of the Initial Experiment

We are considering the difference in performance between each separate layer in an RGB image, compared to the performance when the image is converted to a greyscale image.

Whenever we built the eigenfaces or created the templates, we used images without any extra light. First we analyzed the images when the volunteers looked forward. In Table 1 the EER are given for various cases. Recall that pictures 1-5 were taken without any extra light. This means that the EERs reported in the first line in the table do not take any extra light into consideration. Furthermore does image 6 have extra light from the left, image 7 from the front and image 8 from the right. Each of these images are tested against the templates to find the EER. The next line in the table reflects the performance when only images with extra light are considered, while the last line takes all images into consideration.

Table 1. EER in % for frontal images.

Input	Red	Green	Blue	Grey1	Grey2
3-5	0.0	0.0	0.0	0.0	0.0
6	1.7	4.4	11.4	4.5	3.3
7	1.5	3.3	10.0	2.2	1.8
8	1.4	1.5	3.3	1.5	1.5
6-8	2.7	3.8	8.9	3.3	3.8
3-8	1.8	3.7	7.8	3.6	3.3

In this paper we only report the EER of the various tests. We are interested in a general comparison between the performance of the various color layers and greyscale versions. The EER gives in this case a nice indication and presenting the Detection Error Trade-off (DET) curves for each of the tests would take too much space. As an example we do present in Figure 2 the DET curves for frontal images in case the input equals images 3-8 (related to the last line in Table 1).

In all cases in Table 1 we see that we get the best performance when we take only the red layer for authentication. The numbers in Table 1 already support the idea that the red layer might give better performance than the greyscale image. We have run a test to see which combination of red, green and blue would give the best performance. In this test we used images 3-8 for input. We found that the best achievable EER was equal to 1.67% and was reached for both $0.97 \cdot r + 0.00 \cdot g + 0.03 \cdot b$ and for $0.98 \cdot r + 0.00 \cdot g + 0.02 \cdot b$. In Table 2 we see all combinations of red, green and blue that result in an EER less than 1.9%. Furthermore, regarding the 34 possible combinations that result in an EER of less than 2, we noted that the minimal part red was 85%. Overall we can conclude from these numbers that the red layer provides the largest contribution to the recognition rate.

Table 2. Best achievable EER in % for frontal images and inputs 3-8.

Red	Green	Blue	EER
0.97	0.00	0.03	1.67
0.98	0.00	0.02	1.67
0.96	0.00	0.04	1.82
1.00	0.00	0.00	1.82
0.94	0.01	0.05	1.83
0.98	0.01	0.01	1.84
0.92	0.02	0.06	1.86
0.96	0.01	0.03	1.86
0.94	0.04	0.02	1.87
0.93	0.02	0.05	1.88
0.99	0.00	0.01	1.88

Similar EER computations have been done in the cases where the volunteers looked either to the left (see Table 3) and to the right (see Table 4). When looking at these results we realize some remarkable outcome. In Table 3 we see that in case the volunteer is looking to the left and the extra light source comes from the right (i.e. having extra light on the back of the head of the volunteer), the red layer gives actually the worst performance. For all the other cases the red layer gives again the best performance.

In Table 4 the results can be found when the volunteers face to the right, the direction from which some natural light comes. In this case the results are relatively mixed. Again, the red layer outperforms the 2 other color layers, except in the case when the extra light source is directed toward the back of the head of the volunteers. This is the case in image 6, when the extra light is coming from the left. In that case the best performance comes from the blue layer. Again Grey2 is either better than Grey1 or the results are

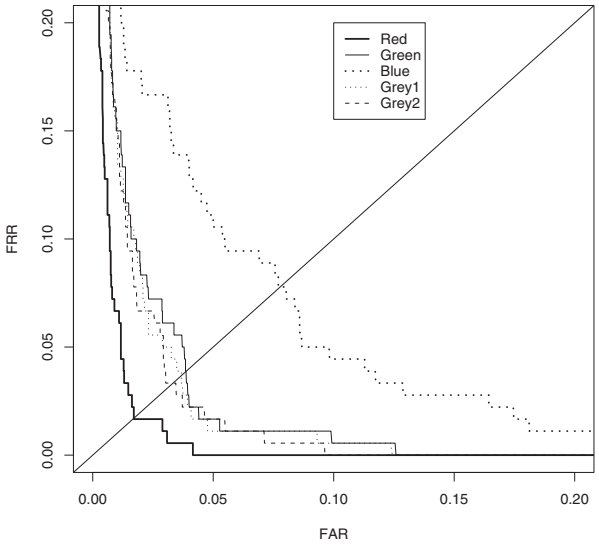


Figure 2. DET curve for input images 3-8 for frontal view

Table 3. EER in % for left facing images.

Input	Red	Green	Blue	Grey1	Grey2
3-5	0.0	0.0	0.3	0.0	0.0
6	5.9	6.7	10.0	8.0	6.7
7	3.3	5.7	8.3	5.2	5.5
8	6.7	4.3	2.2	4.4	5.4
6-8	4.4	6.2	7.8	6.1	6.0
3-8	3.6	4.4	4.7	4.4	4.8

Table 4. EER in % for right facing images.

Input	Red	Green	Blue	Grey1	Grey2
3-5	0.0	0.0	0.4	0.0	0.0
6	5.1	4.9	3.3	4.9	3.3
7	4.3	4.7	7.9	5.1	4.9
8	3.3	6.7	3.3	3.3	3.3
6-8	5.8	5.6	7.8	5.6	4.4
3-8	3.9	4.2	5.3	3.9	3.2

comparable (in case the extra light comes from the front).

6 Data Analysis and Results for the Large Scale Experiment

The data analysis for the larger experiment was done along the lines of the analysis for the initial experiment. A difference was that the camera was a Nikon D70 but the images were again stored in the proprietary NEF format. The cropping was now done based on the location of the eyes in the image. The cropped images were again resized to 100x100 pixels. For notation purposes we will denote the sets of images with no extra light by $\mathbf{N}$ and the sets with light from one of the three directions by $\mathbf{L}$ (for left), $\mathbf{R}$ (for right), and $\mathbf{F}$ (for front). Each of these sets contained 7 images per person, so 7*102 images in total.

The analysis is performed as follows. First we selected the set of images that was to be analyzed and this set was analyzed in two different ways. First with the template based upon images of the set $\mathbf{N}$, next with the template based on the chosen set. Obviously for the set $\mathbf{N}$ these are actually the same. The images that are used to build the PCA are not used again to build the template and both these images are then again not used in the further analysis. The images for the PCA and for the template are furthermore chosen at random from their set. The analysis gave the EER value for that particular setting and each setting was tested 10 times and the average of the resulting EER's was computed. These

values are reported in Table 5. The first two columns in that table report the set that is analyzed (Set) and the set where the PCA images comes from (PCA).

Table 5. EER in % for large scale experiment.

Set	PCA	Red	Green	Blue	Grey1	Grey2
N	**N**	5.5	7.6	12.4	6.9	6.4
L	**N**	6.0	6.1	6.2	6.0	5.9
L	**L**	5.4	5.8	5.4	5.7	5.8
F	**N**	3.7	5.2	7.8	5.3	4.7
F	**F**	3.6	4.7	6.4	4.4	4.4
R	**N**	4.3	5.1	5.7	4.9	4.7
R	**R**	3.9	4.6	5.3	4.6	4.4

In all cases in Table 5 we see that the red layer performs best. The only exception is when the analyzed set and the PCA set are both equal to **L**. In that case all layers and greyscale versions perform more or less the same.

7 Conclusions and Future Research

From the results of the analysis on our databases in the Sections 5 and 6 we see that it might indeed be a good idea to consider the red layer of the RGB color image for face recognition instead of converting the images first to a greyscale. In our analysis we only used PCA in combination with Euclidean distance on the different color layers or greyscale converted images. Future research with different analysis techniques is needed to generalize our results.

The reason why the red layer performs best is not entirely clear. One reason could be that the light used in our experiments had a large peak in the red area in the spectral graph. Another reason could be that in the initial experiment all (except one) of our volunteers are Caucasian with a light, pink skin, so that the red layer of the image could maybe best represent the skin tissue color. This however does not hold for the larger experiment, where there is a mixture of Caucasian and non-Caucasian persons. It is known that skin color has an influence on the performance of face recognition systems [1]. A third factor that could be of influence on the results is the characteristics of the camera. The true reason will most likely be a combination of these three (and maybe more) factors. Another experiment is needed to find the influences of the before mentioned factors on the performance of each of the separate color layers and the combined greyscale images.

From the results in Tables 3 and 4 we see that the performance of the red layer is not the best one in case the extra light is directed toward the hair of the volunteers, so the skin from the face is in the shaded part of the image. This is however a rare situation in a real face recognition system. It however supports the idea that the color of the skin has an influence on the performance when considering the separate color layers. The differences in results for the red layer and the grey2 conversion in Table 5 are not that large, which might be due to the fact that the group of volunteers was a mixture of Caucasian and non-Caucasian people. It could be that a combination of 2 or more layers of the RGB image would give the best recognition performance. In such a case the information that is extracted from the image could be made dependent on the color of the skin. However before such a claim could be proved again more research is needed.

Obviously the results of our experiment leave a lot of open questions, like what is the influence of the light and the camera on the performance, and can we, based on a small set of trial images determine beforehand which (color) information in the image will give the best recognition results. Maybe it is even possible to combine the performances of the three color layers to get an even better overall performance. In this case the color layers are not combined to one greyscale image, but the results after the matching phases in the recognition system are all taken into consideration to make a final decision.

More research needs to be performed to answer questions like "*Does the red layer of an RGB image contain the best information because it resembles the best the color of the skin for Caucasian people?*" or "*What (combination of) color layer(s) performs best in case of non-Caucasian people?*". Furthermore, the fact that we have the best results for the red layer of the image might indicate that using Near Infrared (NIR) lighting and image capturing could also give good results for face recognition. This is again a topic for future research.

Acknowledgments

The authors would like to thank all the volunteers who were willing to participate in our experiments.

References

[1] ATOS Origin. Uk passport service - biometrics enrollment. Technical report, ATOS Origin, May 2005.

[2] D. Blackburn, M. Bone, and P. Phillips. Facial recognition vendor test 2000 -evaluation report. Technical report, NIST, February 2001.

[3] R. Byrd and R. Balaji. Real time 2-d face detection using color ratios and k-mean clustering. In *Proceedings of the 44st Annual Southeast Regional Conference*, pages 644–648. ACM Press, March 2006.

[4] S. Gong, S. McKenna, and J. J. Collins. An investigation into face pose distributions. In *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (FG '96)*, pages 265–270. IEEE, October 1996.

[5] IrfranView. www.irfanview.com, last visited: 21-11-2007.

[6] K. Jia and S. Gong. Multi-modal face image super-resolutions in tensor space. In *Proceedings of IEEE International Conference on Advanced Video and Signal-based Surveillance*, pages 264–269. IEEE, September 2005.

[7] Y.-C. Li and G.-D. Su. Pose discrimination based on cpca-svm in dynamic systems. In *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics*, pages 3970 – 3973. IEEE, August 2006.

[8] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer Professional Computing. Springer, 2nd edition, 2003.

[9] A. Martin, P. Phillips, M. Przybocki, and C. Wilson. An introduction evaluating biometric systems. *Computer*, 33, 2000.

[10] A. Martinez and R. Benavente. The ar face database. Technical Report 24, CVC Technical Report, June 1998.

[11] P. Phillips, P. Flynn, W. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA*, pages 947–954, 2005.

[12] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and M. Bone. Facial recognition vendor test 2002 -evaluation report. Technical report, NIST, March 2003.

[13] P. Phillips, H. Moon, and S. Rizvi. The feret verification testing protocol for face recognition algorithms. Technical Report NISTIR 6281, NIST, October 1998.

[14] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The feret evaluation methodology for face-recognition algorithms. Technical Report NISTIR 6264, NIST, October 1998.

[15] P. Phillips, W. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe. Frvt2006 and ice2006 large-scale results. Technical report, NIST, March 2007.

[16] R-project. www.r-project.org, last visited: 21-11-2007.

[17] M. Rajapakse, J. Tan, and J. C. Rajapakse. Color channel encoding with nmf for face recognition. In *Proceedings of International Conference on Image Processing*, volume 3, pages 2007–2010, 2004.

[18] S. Shan, W. Gao, B. Cao, and D. Zhao. Illumination normalization for robust face recognition against varying lighting conditions. In *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pages 157–164. IEEE, October 2004.

[19] M. Turk and A. Pentland. Face recognition using eigenfaces. *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pages 586–591, June 1991.

[20] X. Wu. Yiq vector quantization in a new color palette architecture. *Image Processing, IEEE Transactions on*, 5(2):321–329, February 1996.

[21] F. Xie, L. Tao, and G. Xu. Estimating illumination parameters in real space with application to image relighting. In *Proceedings of 13th annual ACM international conference on Multimedia*, pages 1039–1040. ACM Press, November 2005.

[22] Yale Face Database B. cvc.yale.edu/projects/yalefacesb/yalefacesb.html, last visited: 21-11-2007.

[23] A. Yip and P. Sinha. Contribution of color to face recognition. *Perception.*, 31(8):995–1003, 2002.

[24] X. Yu and G. Baciu. Face recognition from color images in presence of dynamic orientations and illumination conditions. In *International Conference on Bioinformatics and its Applications*, volume 3072/2004, pages 227–233, July 2004.

[25] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, December 2004.

Part IV
Strengthening Passwords

Chapter 12
Password Generation and Search Space Reduction

Kirsi Helkala and Einar Snekkenes

In *Journal of Computers*, Vol. 4, Issue 7, pp. 663-669, Academy Publisher, Finland, 2009

Password Generation and Search Space Reduction

Kirsi Helkala and Einar Snekkenes
Norwegian Information Security Laboratory, NISlab,
Gjøvik University College, Norway,
Email: {firstname.surname}@hig.no

***Abstract*— It is easy for humans to design passwords that are easily remembered. However, such passwords may have a predictable structure, making exhaustive search feasible. We have divided human-generated passwords into three categories: Non-word passwords, Mixture passwords, and Word passwords; depending on their overall structure. Within these categories, we have analyzed the search-space reduction of several common password sub-structures. From this analysis, we have derived guidelines that yield strong passwords within in each password category. Our results contribute towards the goal of achieving both strong *and* memorable passwords.**

***Index Terms*— Password security, password policy, search space reduction, personnel authentication**

I. Introduction

Passwords are only as strong as the password-designing process. Random passwords can only be produced by random password generators. However, the generated strings might be difficult to remember, especially if someone has many accounts and therefore many different passwords. Humans can easily create memorable passwords, but this also creates the problem that their generation process is guessable, e.g. following structures of certain language [1] or themes [2]. Therefore, human-made passwords are less secure than random passwords.

To help users to generate good passwords, there are guidelines for password creation. However, such guidelines are very general, like those listed in [3], and may not be helpful for all users, given the variety of memorizing techniques. In order to overcome this problem, experts generally recommend [4] a system for evaluating each password against some metric and rejecting the weak ones, rather than mandating a certain number of characters from some character set.

The given guidelines are often based on the use of common knowledge, and not based on scientific computations. Statements such as "Use at least 2 digits, 2 lower case letters, 2 upper case letters, and 2 special characters" might misleadingly guide users into designing passwords with exactly the same number of characters and in the same order as the above statement, such as *12asLK!?*. Such passwords are weaker than the guidance intends, because it reveals a pattern to an adversary. The original meaning was to encourage users to design passwords longer than 8 characters and with characters from all available sets. If the characters were taken randomly from each set, the password would have been quite strong.

In our work, we have computed how much information the adversary gains when the password policy and the generation process are revealed. Based on the findings in [5], [6], we divided human-generated passwords into three categories: Non-word passwords, Mixture passwords, and Word passwords. Non-word passwords are character strings, which do not contain any real words that are found in the dictionary, names, locations etc. However, they can contain letters. Mixture passwords are character strings containing both word and non-word part(s), e.g. *T!today65?* has two non-word parts around the word part in the middle. Word passwords are then strings, which are either pure dictionary words, e.g. *password* or modifications of them e.g. *P@$$W0RD*.

The findings of the information leakage are further used to provide password-generation guidelines for each password category, in such as way that, even if the adversary knows the guidelines, the passwords generated according to these guidelines can be considered as secure.

The remainder of the paper is structured as follows. The analysis is presented in Section II. Section III provides the guidelines for password design. The comparison and discussion of our results and the results of related work is in Section IV. Section V concludes the paper.

II. Analysis of Password Structure

The analyzed cases are shown in Fig. 1. The minimal information which the adversary would gain is the general password policy. In this paper, the basic policy is "Minimum length of 8 characters, maximum length of 14 characters and all visible keyboard keys (except space) are allowed." With a Norwegian keyboard, the number of characters is then 105 and therefore, the maximum password entropy is 94.01 bits, computed as follows

$$\log_2 \sum_{i=8}^{14} 105^i = 94,01 \text{ bits}. \quad (1)$$

This baseline is used when the revealed information is computed with the following formula

$$\mathrm{H}_{Case} = 94,01 - \log_2 \mathrm{C}_{\mathrm{Case}}. \quad (2)$$

The work has received financial support from the Research Council of Norway under grant 158777/530.

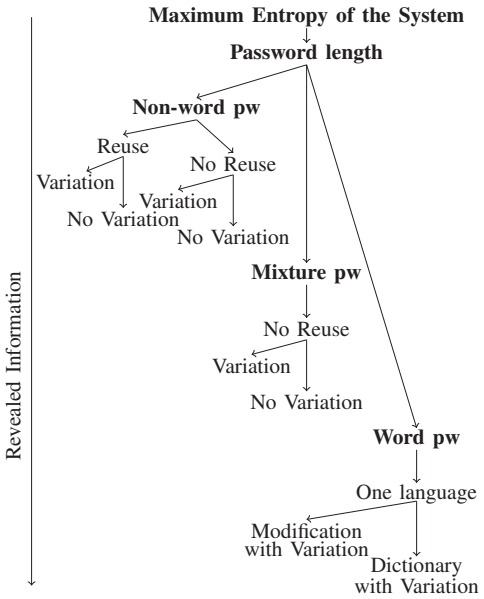


Figure 1. The analysis cases of our study.

We call a password *a good password*, when, given information about the password structure, the revealed information in bits is less than half of the baseline bits, i.e. less than 47 bits.

In the next case, in addition to general policy, the adversary knows the length of the password. The rest of the analysis is divided into three cases: A Non-word, a Mixture and a Word password. A Non-word password does not contain any Norwegian words, a Mixture password contains both Norwegian words or a word and extra characters, and Word passwords contain only words. We use four character sets (digits $|\mathrm{D}| = 10$ (cardinality), lower case $|\mathrm{LC}| = 29$ and upper case letters $|\mathrm{UC}| = 29$, and special characters $|\mathrm{SC}| = 37$). Because we do not have statistics of the most commonly used password characters, we assume a uniform distribution, when selecting characters from each of the above character sets.

A. Knowing the Password Length

In this case, the adversary knows only the system guidelines and length of a password. When only the length l, in addition to the number of the allowed characters cs, is known, the size of the effective search space is

$$\mathrm{C}_{Length} = cs^{l}. \tag{3}$$

The use of Formula 3 gives us following results: With an 8-characters password, the adversary knows 40.3 bits and with a 14 characters long password, the adversary knows 0.01 bits. If the adversary learns that a password was actually 7 characters long, he would have learned 47 bits. This suggests that passwords shorter than 8 characters are *not good*.

B. Human Design Passwords

Password strength depends on the password-design process. Maximum password strength is achieved when each password character is drawn independently and a uniform distribution is used. Usually, this is not the case when people design their own passwords.

Human-designed passwords can be Non-word, Mixture or Word passwords. Non-word password do not contain sub-strings that can be found in a dictionary e.g *NT*-Ke0*, Mixture passwords contain some words and extra characters e.g *4seasons1year* and Word passwords contain only words e.g *SkiingIsTheBestIKnow*. We show which design processes within each category provide good passwords and which do not. The analysis shows that the main criterion for designing good passwords is to vary the characters used and the placements of characters within each password-design session.

Non-word passwords. The analysis is divided into four sub cases.

- NW**rapu**: **r**eused characters are **a**llowed and the **p**attern of character placement in a password is **u**nknown.
- NW**rapk**: **r**eused characters are **a**llowed and the **p**attern of character placement in a password is **k**nown.
- NW**rdpu**: **r**eused characters are **d**enied and the **p**attern of character placement in a password is **u**nknown.
- NW**rdpk**: **r**eused characters are **d**enied and the **p**attern of character placement in a password is **k**nown.

We assume that if a person always designs passwords with the same structure, the pattern of the password is known.

Computations. The size of the Non-word password search space is

$$\mathrm{C}_{\mathrm{NonWord}} = f_1 g_1 \times f_2 g_2 \times f_3 g_3 \times f_4 g_4 - \mathrm{W}, \tag{4}$$

where functions g_i, computed with (5)-(8), give the cardinality of each character set used in a password and functions f_j, computed with (9)-(12), give the number of all possible combinations of character placement for each set in a password of length l. W stand for the number of possible words within letter combinations. These combinations are subtracted from the total number of the combinations, because Non-word passwords do not contain words. In these formulae, uc stands for the number of upper case letters, lc for lower case letters, d for digits, and sc for special characters.

The cardinality functions g_i are as follows

$$g_1(uc) = \begin{cases} 29^{uc}, & \text{reuse allowed} \\ \frac{29!}{(29-uc)!}, & \text{reuse denied} \end{cases} \tag{5}$$

$$g_2(lc) = \begin{cases} 29^{lc}, & \text{reuse allowed} \\ \frac{29!}{(29-lc)!}, & \text{reuse denied} \end{cases} \tag{6}$$

$$g_3(d) = \begin{cases} 10^d, & \text{reuse allowed} \\ \frac{10!}{(10-d)!}, & \text{reuse denied} \end{cases} \quad (7)$$

$$g_4(sc) = \begin{cases} 37^{sc}, & \text{reuse allowed} \\ \frac{37!}{(37-sc)!}, & \text{reuse denied} \end{cases} \quad (8)$$

Equations (9)-(12) are character-placement combination functions. These functions will get a value 1 if the pattern of the password is known. In other words, the adversary knows which characters in a password are digits, upper case letters, etc. When the pattern is unknown, the placement combinations are computed as follows

$$f_1(l, uc) = \binom{l}{uc} \quad (9)$$

$$f_2(l, uc, lc) = \binom{l-uc}{lc} \quad (10)$$

$$f_3(l, uc, lc, d) = \binom{l-uc-lc}{d} \quad (11)$$

$$f_4(l, uc, lc, d, sc) = \binom{l-uc-lc-d}{sc} \quad (12)$$

The number of passwords containing single or multiple words with a length of $(uc + lc)$ letters written both forwards and in reverse transformation, W, is computed with (13). We have simplified the subtraction, considering only the passwords containing words formed from all the letters in a particular password. The number of words, aw, shown in Table III, were provided by the Norwegian Text Laboratory.

$$W = [l - (uc + lc) + 1] \times 2aw \times f_3 \times g_3 \times f_4 \times g_4. \quad (13)$$

Example. As an example of the use of formulae, we show a situation in which a 10- character password ($l = 10$) contains 1 digit ($d = 1$), 1 lower case letter ($lc = 1$), 4 upper case letters ($uc = 4$) and 4 special characters ($sc = 4$). The reuse of the characters is allowed and the pattern is unknown. The number of words with a length of 5 letters is taken from Table III. Because the reuse of characters is allowed, we obtain the following cardinalities for the sets

$$\begin{aligned} g_1(4) &= 29^4 \\ g_2(1) &= 29 \\ g_3(1) &= 10 \\ g_4(4) &= 37^4. \end{aligned} \quad (14)$$

The combinations of the character placements will then be

$$\begin{aligned} f_1(10, 4) &= \tbinom{10}{4} = 210 \\ f_2(10, 4, 1) &= \tbinom{10-4}{1} = 6 \\ f_3(10, 4, 1, 1) &= \tbinom{10-4-1}{1} = 5 \\ f_4(10, 4, 1, 1, 4) &= \tbinom{10-4-1-1}{1} = 1. \end{aligned} \quad (15)$$

These will then yield the search space size of

$$\begin{aligned} \mathrm{C} \quad &= 210 \times 29^4 \times 6 \times 29 \times 5 \times 10 \times 37^4 \\ &\quad -[10 - (4 + 1) + 1] \times 2 \times 20767 \times 5 \times 10 \times 37^4 \\ &= 2.42 \times 10^{18} \end{aligned} \quad (16)$$

which will reveal information of

$$\mathrm{H} = 94,01 - \log_2 2.42 \times 10^{18} = 32.9 \text{ bits}. \quad (17)$$

TABLE I.
REVEALED KNOWLEDGE OF NON-WORD PASSWORDS. THE USED SETS ARE DIGITS (D), UPPER CASE (UC) AND LOWER CASE LETTERS (LC), AND SPECIAL CHARACTERS (SC). THE NUMBERS IN EACH CHARACTER SET COLUMN GIVE THE NUMBER OF CHARACTERS USED FROM EACH SET IN THE PASSWORD-DESIGN PROCESS. GOOD PASSWORDS ARE IN BOLD.

Pw L	Nr D	Nr LC	Nr UC	Nr SC	NW-rapu Bits	NW-rapk Bits	NW-rdpu Bits	NW-rdpk Bits
10	10	0	0	0	60.8	60.8	72.2	72.2
	9	1	0	0	55.9	59.3	64.0	67.4
	8	1	1	0	51.2	57.7	57.0	63.5
	0	0	0	10	**41.9**	**41.9**	**43.8**	**43.8**
	0	0	1	9	**38.9**	**42.3**	**40.5**	**43.8**
	0	1	1	8	**36.1**	**42.6**	**37.3**	**43.8**
	0	0	5	5	**35.7**	**43.7**	**36.6**	**44.6**
	1	4	4	1	**34.0**	**46.6**	**34.6**	47.2
	1	1	4	4	**32.9**	**45.6**	**33.5**	**46.1**
	0	2	3	5	**32.4**	**43.7**	**33.0**	**44.3**
8	8	0	0	0	67.4	67.4	73.2	73.2
	6	1	0	1	58.2	64.0	60.9	66.7
	0	0	0	8	52.3	52.3	53.5	53.5
	1	1	0	6	48.8	54.6	49.4	55.2
	0	0	4	4	47.6	53.7	48.2	54.3
	1	1	1	5	**46.5**	54.9	**46.9**	55.3
	2	2	2	2	**46.2**	57.5	**46.5**	57.8
	0	1	3	4	**45.6**	53.7	**46.0**	54.1
	1	1	3	3	**45.5**	55.6	**45.8**	55.9
	0	3	3	2	**45.3**	54.4	**45.7**	54.8
	1	3	2	2	**45.3**	56.0	**45.5**	56.2
	0	2	3	3	**45.0**	54.1	**45.3**	54.4

Other examples of the information revealed in each of four cases with a password length of 8 and 10 are shown in Table I.

Results. In summary, passwords are *not good* if they consist only of digits, or of digits and letters only either from the lower case letter set or the upper case letter set. The best passwords contain some special characters and characters from other sets, so that the number of digits is as low as possible.

When concentrating only on strong passwords (the bold entropies in Table I, we find the following. Comparison of the "reuse of characters allowed" -columns (*NWrapu* and *NWrapk*) to "reuse of characters denied" -columns (*NWrdpu* and *NWrdpk*), shows that the revealed information is rather similar. When characters are from several sets and the most of them are not digits the difference between cases gets smaller, and so do the actual revealed information entropies.

The difference in bits is significant, when comparing "pattern-unknown" columns (*NWrapu* and *NWrdpu*) with "pattern-known" columns (*NWrapk* and *NWrdpk*). This finding strongly supports the need to change the character pattern in each password-design session. According to Table I, it is possible to design a *good* Non-word password of length of 8 characters.

Mixture passwords. Mixture passwords contain both a word and a non-word component. The following analysis is divided into two sub-cases with three different possibilities for the extra character set comprising either a set of

digits, special characters, or digit and special characters. Here, we consider only those cases where the reuse of characters is denied.

- M**pu**: the placement **p**attern of the word(s) and extra character(s) is **u**nknown.
- M**pk**: the placement **p**attern of the word(s) and extra character(s) is **k**nown.

In order to compute the worst case scenarios, we only consider dictionary words as words, not their modification. Their modification is discussed further in the section below on Word passwords.

Computations. The size of the Mixture password search space is

$$C_{Mixture} = c_{wo}c_{pe}(\prod_{p=1}^{w} aw_p(lw_p))\frac{\mathrm{es}!}{(\mathrm{es}-n)!}, \quad (18)$$

where $aw(lw)$ is the number of lw -length words, w the number of words, es the size of the extra character set and n the number of extra characters used in non-word parts of password. The combination of different word orders c_{wo} is computed with (19), where w_{sl} is number of same-length words. The combinations of extra character places between words c_{pe} was noted, so as to follow the numbers in Pascal's Triangle, which is then used to compute (20). Again, if the pattern is known, then (19) and (20) are 1. If the pattern is unknown, then the formulae are as follows

$$c_{wo} = \frac{w!}{w_{sl}!} \quad (19)$$

$$c_{pe} = \sum_{i=1}^{k} \binom{n-1}{i-1}\binom{w+1}{i}, \quad (20)$$

where $n \geq 1$ and $k = \min(n, (w+1))$.

Example. As an example of the use of the formulae, we show a case where a password contains 2 words ($w = 2$), both with a length of 3 letters ($w_{sl} = 2$), and 4 digits ($n = 4$, $es = 10$). The possible word order combinations are

$$c_{wo} = \frac{2!}{2!} = 1 \quad (21)$$

and the possible placement combinations among digits and words are

$$c_{pe} = \sum_{i=1}^{3} \binom{4-1}{i-1}\binom{2+1}{i} = 15. \quad (22)$$

These yields the following search space size

$$C = 15 \times (3048^2)\frac{10!}{(10-4)!} = 7.02 \times 10^{11}, \quad (23)$$

which leads to revealed information of

$$\mathrm{H} = 94,01 - \log_2(7.02 \times 10^{11}) = 54.7 \text{ bits}. \quad (24)$$

Some other examples of the revealed information of Mixture passwords with a length of 8 and 10 characters are shown in Table II.

Results. The revealed information entropies in Table II show that a *good* Mixture password cannot be shorter than 10 characters. Furthermore, even with 10 characters, only a small number of the passwords can really be considered as *good*.

TABLE II.
REVEALED INFORMATION OF MIXTURE PASSWORDS. THE USED SETS ARE DIGITS, D, SPECIAL CHARACTERS, SC, AND COMBINATION OF DIGITS AND SPECIAL CHARACTERS, DSC. EC STANDS FOR EXTRA CHARACTERS. GOOD PASSWORDS ARE IN BOLD.

Pw L	Word L+ Nr EC	Mpu D	Mpu SC	Mpu DSC	Mpk D	Mpk SC	Mpk DSC
10	9 + 1	74.2	72.3	72.0	75.2	73.3	73.0
	6,3 + 1	61.7	59.8	59.5	64.3	62.4	62.1
	6 + 4	64.6	56.3	54.8	66.9	58.6	57.2
	5,4 + 1	60.3	58.4	58.1	62.9	61.0	60.7
	5,3 + 2	58.0	54.1	53.4	61.6	57.7	57.0
	4,4 + 2	58.0	54.2	53.5	60.6	56.7	56.0
	4,3 + 3	55.2	49.2	48.1	59.5	53.5	52.4
	3,3,3 + 1	54.0	52.1	51.7	56.0	54.1	53.7
	3,3 + 4	54.7	**46.4**	**44.9**	58.6	50.3	48.8
	3 + 7	60.2	**43.8**	**41.2**	63.2	**46.8**	**44.2**
8	7 + 1	74.5	72.6	72.3	75.5	73.6	73.3
	6 + 2	71.1	67.2	66.5	72.7	68.8	68.1
	5 + 3	68.2	62.2	61.1	70.2	64.2	63.1
	4,3 + 1	63.1	61.2	60.9	65.7	63.8	63.4
	4 + 4	65.9	57.6	56.2	68.3	60.0	58.5
	3,3 + 2	61.8	57.9	57.2	64.4	60.5	59.8
	3 + 5	65.0	54.2	52.4	67.6	56.8	55.0

The difference in revealed information between "unknown" and "known pattern" is a couple of bits based on a comparison of the columns *Mpu* and *Mpk*. However, the amount of revealed information in the *D*, *SC*, and *DSC* -columns show that the use of extra characters from the larger set, yield a much stronger password. The impact is greater, when most of the password characters are extra ones and do not belong to a word-part.

From the above, it can be concluded that the best Mixture passwords consist of a short word (or couple of short words) and many extra characters from a large character set. The short word in our computation was noticed to be a word less than half of the length of the complete password. For instance, for a password of 10 characters, the short words can have a length of 3 and 4 characters.

Word passwords. By pure Word passwords, we mean passwords which contain dictionary words only. The password either contains only one word with the length of the password itself, or several shorter words, with the total length of the words constituting the total length of the password. The analysis of words from one language is divided into two sub-cases.

- W**mod**: the **mod**ified words are used and the placement pattern of the words is unknown.
- W**dic**: the **dic**tionary words are used and the placement pattern of the words is unknown.

It has been shown above that a variation of character placements in each password-design session makes passwords more secure. In the case of Word passwords, the variation is done by using different word lengths. This means that the effect of variation is smaller than in other cases, because there are fewer words than characters in a

TABLE III.
NUMBER OF NORWEGIAN WORDS. THE STATISTICS WERE OBTAINED FROM THE OSLO CORPUS COLLECTION BY NORWEGIAN TEXT LABORATORY, TEKSTLABORATORIET ILN, OSLO, NORWAY.

Word length, l	Number of words, aw
3	3048
4	11145
5	20767
6	29043
7	36590
8	42805
9	45762
10	44956

password. However, we consider that even if the password structure were same, a change in words makes it possible to use different word lengths. For example, *ABussIsWhite* has a different word-length structure than a password with a similar theme password such as *TheTruckWasBrown*. Therefore, we only consider cases in which the word placement pattern in a password is unknown.

In our work, we only compute revealed information entropies when only one language is used, and we decided to use the statistics for Norwegian words. However, the formulae provided are also suitable for other languages.

Computations. The size of the Word password search space is

$$C_{Word} = c_{wo} \prod_{p=1}^{w} aw_p(lw_p), \tag{25}$$

where c_{wo} is the order combination of different words (computed with (19)), w is the number of words, and $aw_p(lw_p)$ is the amount of lw_p -length words. Table III shows the number of Norwegian words with a length of 3-10 letters.

Results. The results for the 8, 10, and 12-character passwords are shown in Table IV. In the *Wmod* case, both forward and reverse transformations are taken into account. The modifications are: all letters are lower case, all letters are upper case, only the first letter is upper case, only the last letter is upper case, the consonants are upper case and the vowels upper case.

The results in Table IV show that the use of original dictionary words does not make passwords secure, especially, if only one or two words are used. The best approach is to use several short words, with different lengths, in one password. Here also, the length of the short word is less than half of the original password length. The use of modified words makes the passwords stronger. However, if the modification is always done in the same manner, the use of modification reduces the password space as much as use of original dictionary words.

It is possible to design strong Word passwords, but not with pure dictionary words. A *good* Word password need to consist of digits, upper case and lower case letters, and special characters. The substitution and rotation should be done differently each time and also differently in each password. A *good* Word password should have at least 12 characters and consists of several short, modified words.

TABLE IV.
REVEALED INFORMATION OF WORD PASSWORDS. GOOD PASSWORDS ARE IN BOLD.

Pw L	Word L	Wmod bits	Wdic bits
12	12	75.1	78.7
	9,3	58.8	66.0
	8,4	57.0	64.2
	7,5	56.3	63.5
	6,6	57.2	64.4
	6,3,3	**43.7**	54.5
	5,4,3	**41.3**	52.1
	4,4,4	**42.9**	53.7
	3,3,3,3	**33.4**	47.7
10	10	75.0	78.6
	7,3	59.1	66.3
	6,4	57.6	64.7
	5,5	58.2	65.3
	4,3,3	**45.1**	55.8
8	8	75.0	78.6
	5,3	59.9	67.1
	4,4	60.0	67.1

The modified Word passwords look like Mixture passwords, but, because they are constructed from dictionary words, the underlying word pattern makes them weaker than Mixture-words. We do not have statistics specifying which letters are modified and by which other character, but it can be assumed that substitution by people follows the certain pattern.

The number of dictionary entries is small compared to the total size of the password search space. However, the actual size of word-sets used in passwords might be even smaller. People have tendency to use theme words [2], [7] such as name of the sport teams, food, and animals. The size of such themes is very small and the use of words only from one theme makes password design process very weak.

III. PASSWORD FORMATION GUIDELINES

A good password is complex, but nonetheless easy to remember [8]. The password policy should be such that it combines individual password design processes, while helping users to generate secure passwords with their own methods. In Section II, we showed that *good passwords* can be created in each category, if there are enough variations in pattern and character. Variations provide good defences against attacks based on language structures e.g. fast dictionary attack in [1].

In order to design *good passwords*, we propose the following guidelines.

Non-word password design.

1) A password should be longer than 8 characters.
2) Use characters from all character sets, so that more characters come from the large character sets than from the small sets.
3) Vary the number of characters from each set in each construction session.
4) Vary the patterns of character placement.

Mixture password design.

1) A password should be longer than 10 characters.
2) Use either one short, modified word and many extra characters or several short (not the same length), modified words and a few extra characters from large character set.
3) Avoid using same theme.

Word password design.

1) A password should have more than 12 characters.
2) Use many short and modified words.
3) Avoid using same theme.
4) Use variation when modifying.
5) Use different languages and language combination when designing a new password.

Note that the length of a short word is less than half of the length of the password.

IV. Related Work and Discussion

When designing passwords and password policies, users should also develop an understanding that passwords are not only vulnerable against brute-force attack, but also against much more sophisticated attacks. A common denominator of the latter is *password search space reduction*. After reduction, the attacks, be they brute-force or dictionary attacks in smaller search spaces, become faster and therefore constitute a far more serious threat.

In our study, the reduction of password search space is computed from the adversary's acquired knowledge on the password policy and the password-generation process. More sophisticated methods for password search space reduction are presented in [1], [9]–[11].

Trostle [9] describe timing attacks against the trusted path mechanism. Only a few trials were needed to obtain the length of the password with the first attack type. The second attack continued to obtain leakage and was able to reduce the strength of a password by 2-3 bits per character.

In a study of Song et al. [10], the use of keystroke latency information in timing attacks toward passwords in the Secure Shell was analyzed. Typing patterns were estimated and, with the help of latency information, the strength of a password was reduced by 1.2 bits per character pair.

In [1] Narayanan and Shmatikov reduced the password search space by using Markov modelling techniques of natural language processing. They claim that the distribution of letters in easily remembered passwords is similar to the distribution of letters in the users' native language. Based on this concept, they combined an algorithm which enables a fast dictionary attack. First Markov filters are used to reduce the size of the password search space, and then, the remaining search space is efficiently enumerated and in the final stage, time-space trade off techniques are used to conduct a fast dictionary attack.

Markov filters and English language structures were also in use when Zhuang et al. [11] presented their keyboard acoustic emanation attack. They recorded 10 minutes of English text and were able to recover 96% of the typed characters.

In the NIST Special Publication on Electronic Authentication Guidelines [12], the use of three different password policies were evaluated and the minimum password guessing entropies, versus password length for each policy, were estimated. Compared to NIST evaluation, we used the password analysis based on textual meta-information, and went one step deeper, by studying human password design processes.

More password policies and their relationship to user memorability, password entropy and password change frequency were simulated and analyzed by Shay et al. [7]. They noted, as had many before [13], [14], that if a password policy does not require sufficiently complex passwords, users' passwords are in danger of being cracked. If a policy requires overly complex passwords, users may have problems to recall them and therefore may write them down.

Password generation process and structure were studied in [5], [6]. MySpace phishing attack analysis [6] showed that 65% of passwords were 8 characters or less long and 81% of passwords were alphanumeric, which, in most cases, contained lower case letters with a single digit at the end. In one third of the cases, the digit was 1. Brown et al. in [5] found that 65% of passwords were generated from information relating to the user himself. Almost one third of these were names. The next largest groups were dates, and ID and phone numbers. Three quarters of the passwords contained the full information. There was only a modification of the information in less than 5% of the passwords. These findings suggest that users generate passwords with minimum length and structure and even the content is familiar. The users of these studies cited above, would benefit from our password policies.

Based on our analysis, we are able to provide concrete guidelines for password policy construction in each password category: Non-Word, Mixture and Word password. This allows users to develop their own password style and still create memorable passwords, while keeping the structure of passwords complex enough to ensure security. Passwords constructed according to guidelines are strong, even if these guidelines are available to the adversary.

An educational tool [15] based on the findings and guidelines of this paper, has been made to help users measure the quality of their password and also to design stronger passwords.

V. Conclusion

People *are* able to design memorable passwords, especially if they can use their own password-designing processes. These processes can produce passwords which we have divided further into three categories: Non-word passwords, Mixture passwords, and Word passwords. Within these categories, we have provided formulae and

computations of effective password space for many different password policies. Based on this analysis, we have compared how effective password space is affected by policy decisions. We have computed how valuable a knowledge of password policy is for an adversary, in terms of reduced password entropy and consequently, a reduction in the necessary search space. Based on computations of typical password sizes (8-14 characters), we have provided guidelines for password policy construction. These guidelines identify policy statements that help to reduce the loss of password entropy.

ACKNOWLEDGEMENTS

The authors are grateful to the anonymous reviewers for their careful reading and valuable feedback. This work is supported by the Research Council of Norway, grant 158777, Authentication in a health service context.

REFERENCES

[1] A. Narayanan and V. Shmatikov, "Fast dictionary attacks on passwords using time-space tradeoff," in *Proc. of 12th ACM conference on Computer and communications security*, 2005, pp. 364–372.

[2] F. Monrose and M. K. Reiter, "Graphical passwords," in *Security and Usability: Designing Secure Systems that People Can Use*. O'Reilly, 2005, ch. 9, pp. 161–179.

[3] S. Gaw and E. W. Felten, "Password management strategies for online accounts," in *Proc. of Symbosium On Usable Privacy and Security*, 2006, pp. 44–55.

[4] E. Gehringer, "Choosing passwords: security and human factors," in *Proc. of International Symposium on Technology and Society*, 2002, pp. 369–373.

[5] A. Brown, E. Bracken, S. Zoccoli, and K. Douglas, "Generating and remembering passwords," *Applied Cognitive Psychology*, vol. 18, pp. 641–651, 2004.

[6] B. Schneier, "Crypto-gram newletter: Real-world passwords," http://www.schneier.com/crypto-gram-0612.html, December 2006.

[7] R. Shay, A. Bhargav-Spantzel, and E. Bertino, "Password policy simulation and analysis," in *Proc. of ACM workshop on Digital identity management*, 2007, pp. 1–10.

[8] P. Cisar and S. Cisar, "Password - a form of authentication," in *Proc. of 5th International Symposium on Intelligent Systems and Informatics*, 2007, pp. 29–32.

[9] J. Trostle, "Timing attacks against trusted path," in *Proc. of IEEE Symposium on Security and Privacy*, 1998, pp. 125–134.

[10] D. X. Song, D. Wagner, and X. Tian, "Timing analysis of keystrokes and timing attacks on ssh," in *Proc. of 10th conference on USENIX Security Symposium*, vol. 10, 2001.

[11] L. Zhuang, F. Zhou, and J. Tygar, "Keyboard acoustic emanations revisited," in *Proc. of 12th ACM conference on Computer and communications security*, 2005, pp. 373–382.

[12] W. E. Burr, D. F. Dodson, and W. T. Polk, *Information Security: Electronic Authentication Guideline*. NIST Special Publication 800-63 Version 1.0.2, 2006.

[13] C. Kuo, S. Romanosky, and L. Cranor, "Human selection of mnemonic phrase-based passwords," in *Proc. of 2nd symposium on Usable privacy and security*. ACM Press, 2006, pp. 67–78.

[14] M. Sasse, S. Brostoff, and D. Weirich, "Transforming the "weakest link" - human/computer interaction approach to usable and effective security," *BT Technol*, vol. 19, no. 19, pp. 122–131, 2001.

[15] K. Helkala, "An educational tool for password quality measurements," in *Proc. of Norwegian Information Security Conferences (Norsk Informasjonssikkerhetskonferanse, NISK)*, November 2008, pp. 69–80.

BIOGRAPHIES

Kirsi Helkala is a doctoral candidate in information security at Gjvik University College, Norway. Her research topic is personnel authentication.

Einar Snekkenes is professor of information security at Gjvik University College, Norway. He has published on such topics as cryptographic protocol analysis, privacy, social engineering, biometrics, authentication, testing and smart card side channel attacks. His research interests include information security and risk analysis.

Chapter 13
An Educational Tool for Password Quality Measurements

Kirsi Helkala

In *Proceedings of Norwegian Information Security Conference (Norsk Informasjonssikkerhetkonferanse) (NISK2008)* , pp. 69-80, Tapir Akademisk Forlag, 2008

Chapter 14
Password Education Based on Guidelines Tailored to Different Password Categories

Kirsi Helkala

Submitted for publication 2010

Password Education Based on Guidelines Tailored to Different Password Categories

Kirsi Helkala
Norwegian Information Security Laboratory, NISlab,
Gjøvik University College, Norway,
Email: {firstname.surname}@hig.no

Abstract— **General password policies do not guarantee that passwords fulfilling the requirement are good enough. The policies have a tendency to be too broad to be useful for all users. Different users have different designing processes based on what kind of passwords they most easily remember. Users are also often left to generate passwords on their own without any training. In our study we measured the effect of education on the strength of a password. In order to help users to create good passwords, we divided passwords into three password categories: Word password, Mixture password and Non-word password. For each category different password generation guidelines were taught. Participants had access to the password quality measurement tool, which not only measured the strength of the password but also guided students in the generation process. It was shown that education had a positive effect and that passwords became stronger right after the education. The most important result was that a password structure got changed as the variation of structures increased and different structure types were more evenly distributed. However, after half a year without reminders or education repetition, most of the positive effect was lost. While password structures still differed, they had become less complex as participants had given up using special characters.**

Index Terms— **Password security, education, personnel authentication**

I. Introduction

Passwords have existed a long time and are commonly in use. Therefore it is common to think that everybody knows how to create, store and manage passwords without education. However, research shows that this is not the case [6], [9], [11].

When a person generates a password, there is often an easily guessable structure behind. One of the most common password structure is a dictionary word ending with couple of digits [1], [12]. Password cracker tools such as "John the Ripper" start guessing by using pure dictionary words or digits in rising order [13]. In order to avoid this problem, password policies [2], [7], [8] are defined to guide people to generate strong passwords. However, even if a system uses password policy, stating which character sets to use and what the minimum length of a password is, passwords might end up being predictable. Often characters are used in exactly the same order as given in the instructions, and most of the passwords are only as long as the minimum password length. Based on predictable password structures, an adversary can carry out targeted attacks against the system with a greater probability of success.

From the user perspective the general instructions are often too broad to be useful. One remembers best meaningful passwords, some like numbers and special characters while others recall best patterns from the keyboard, etc. The general instructions such as "Minimum length of 8, use all character sets" do not support all users in their password generation process. Users have a tendency to remember own-generated passwords better computer-generated ones [15]. Therefore password policies should be such that all users would benefit from them and still be able to generate password with their own style. Sasse et. al. [11] points out that the users should be educated to design stronger passwords and they should be given time and tools for the password generation process. In our study, we motivated users to design strong passwords by allowing them to use the designing techniques that are best suited for them.

We encouraged students to design passwords within three different password categories: Non-word passwords, Mixture passwords, and Word passwords. For each category different password generation guidelines were taught. We gave password education for two groups. One group got a classical classroom lecture and the other did a home study based on class room teaching lecture notes. Both groups were given an access to the password quality measurement tool. The tool uses a questionnaire to derive some simple structures of the passwords. The structure information is then used to compute the quality score of the password. The computations and password policies based on the search space reduction computations are found in [4], and the tool itself is presented in [5].

The remainder of this paper is structured as follows. Section II briefly explains the experiment, Section III summaries the educational part of the study and results are shown in Section IV. Discussion about the study and future work is done in Section V and Section VI concludes the study.

II. Experiment

The study was done with Information Security Bachelor students in their first half semester in 2008 and 2009. Students were taught how to design strong password within

The work has received financial support from the Research Council of Norway under grant 158777/530.

three different password categories: Non-word passwords, Mixture passwords, and Word passwords. However, the teaching method was different. When the first group (2008) was taught by using normal class teaching containing lecturing and free question and discussion about the topic, the second group (2009) did an individual home study based on the same lecture notes. Both groups were asked to design a new password after this.

The password categorization was new to students. Therefore, verbal instruction was given to both groups. After the instruction to the password categories, students were also explained how to use the tool provided to measure the password strength. After these two small sessions, students got the first questionnaire. The questionnaire consists of the questions asked by the tool, see Appendix, and also explanation part where students explained how they had constructed the password. With this questionnaire we measured the strength of the current passwords and gained information of the generation process. The education part followed right after the students had delivered their first questionnaire. The second questionnaire was carried out one week after. It was similar to the first one, but now the students were asked to use a password they had designed based on previous week teaching session. The third questionnaire was carried out again a week after from the second one. Now only designed password was asked in order to see how well they were remembered.

In addition to these, we asked the home study group fill the questionnaire again after half a year from the education. This fourth questionnaire was identical with the first questionnaire.

III. Teaching Users to Design Good Passwords

A. Password Categories

Passwords can be divided into three categories: Non-word passwords, Mixture passwords, and Word passwords. Non-word passwords are character strings, which do not contain any words with any writing styles. However, they can contain letters. Mixture passwords are character strings containing both a word and a non-word part(s), e.g. "T!today65?" has two non-word parts ("T!" and "65?") around the word part ("today"). Word passwords are strings which are either pure dictionary words, e.g. "password" or readable modifications of them e.g. "P@$$W0RD". In our study, we considered modifications shown in Table I as single words. Several examples of the passwords within each category were given to students.

B. A Tool for Password Quality Checking

The tool was given to the students in a form of an Excel-sheet and it was written in Norwegian. Questions of the tool in English are show in Table II. Because passwords can contain both a non-word and a word part, the questionnaire was divided accordingly. The main challenge for student was to understand the separation of these two parts. The strength of the password was not correct if the questionnaire sheet was filled in wrongly. Therefore the instructor's help was important when the tool is used the first time.

TABLE I.
Considered as a single word

Definition	Example
Original dictionary word:	library
Compound dictionary word:	password
Reverse writing:	library → yrarbil
Modifications	
with uppercase letters:	library → LiBRaRy
with digits:	library → l1brary
with special characters:	library → l!br@ry
with uppercases and digits:	library → L1brary
with uppercases and special characters:	library → Libr@ry
with digits and special characters:	library → l1br@ry
with all sets:	library → L1br@ry

TABLE II.
Password quality questions

Whole password	How long is your password?
Word -part	How many words does your password contain?
	What languages are the words based on?
	How many letters does the first word contain?
	How many letters does the nth word contain?
	How many uppercase letters are there?
	How many digits letters are there?
	How many special characters are there?
Non-word -part	How many upper case letters are there?
	How many lower case letters are there?
	How many digits are there?
	How many special characters are there?
	How many reused characters are there?

In [4] we have analysed the information leakage of passwords when the adversary has an access to the questionnaire answers. Based on these computations the tool used in this study was build. The students were allowed to use the tools freely after the first instruction session. And they were told that a password which scores more than 735 is *a good password* and a password achieving more points than 875 is called *a strong password*. The score points levels are based findings in both in [4] and in [3].

C. Password Creation

The students in this study were taught how to create good and strong passwords among each category separately. The guidelines for the password design are again based on the findings in [4].

Word password. Students were taught that a good word password should be longer than 12 characters and it should contain modified words shorter than the half of the length of the actual password. The best modification score will be acheived when all character sets are taken into use when modifing. The words would preferably come from different themes (such as sport and music) and different languages. An example of a good Word password is shown in Table III. The guidelines for Word passwords are following

1) A password should have more than 12 characters.
2) Use many short and modified words.
3) Avoid using same theme.

TABLE III.
EXAMPLES OF GOOD PASSWORDS IN EACH CATEGORY

Password	Description	Score
Word: $K@1#y@H0$f@R	Based on “Skal ya hos far”, containing 13 characters, 4 words: mod with 3 ul, 2 d, 6 sc	749
Mixture: EeSn#&S0l3!	Based on “Jeg elsker sn og sol”, containing 11 characters, 2 words: mod with 2 uc, 1 d, 1 sc, non-word: 1 lc, 1 uc, 1 d, 2 sc	808
Non-word: 7(3-9f)>K	Based on a math formula, containing 9 characters, non-word: characters from all sets	945

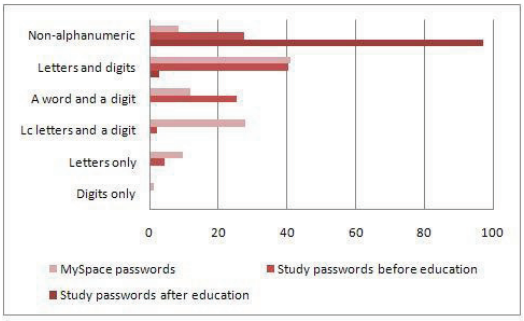


Figure 1. Password structures with common categorization

4) Use variation when modifying.
5) Use different languages and language combination when designing a new password.

Mixture password design. Students were given similar instructions as given for the Word passwords considering the word-part. The strength of the password can be increased when all character sets are taken into use in non-word parts. Table III shows an example of a good Mixture password. The guidelines for Mixture passwords are following

1) A password should be longer than 10 characters.
2) Use either one short, modified word and many extra characters or several short (not the same length), modified words and a few extra characters from large character set.
3) Avoid using same theme.

Non-word passwords. The best Non-word passwords are random passwords containing characters from all character set. However, they might be hard to remember. Therefore, students were shown some examples how to generate random-looking passwords in a way that they are easier to remember, etc. mnemonic passwords [10], [14]. Table III shows an example of a strong Non-word password. The guidelines for Non-word passwords are following

1) A password should be longer than 8 characters.
2) Use characters from all character sets, so that more characters come from the large character sets than from the small sets.
3) Vary the number of characters from each set in each construction session.
4) Vary the patterns of character placement.

IV. RESULTS

A. Generation Process

Passwords are often categorized base on what kind of characters they consist of: only digits, only letters, alphanumeric, non-alphanumeric, etc. In the data collected from phishing attack on MySpace 2006 [12], the character mix in passwords were following: 1,3% contained only digits, 9,6% contained only letters, 81% were alphanumeric and the rest 8,3% were non-alphanumeric. Taken a deeper look to the data, among the alphanumeric passwords, a structure pattern was found. 28% of passwords contained lowercase letters having a single digit in the end. 12% contained a single dictionary word having a single digit in the end. 3,8 % of passwords were a single dictionary word.

With similar categorization, the passwords in our study before the education were as follows. 4,3% of passwords contained only letters (being also pure words), 25,5% consisted of a single word and digits and in total 68,1% were alphanumeric. 27,7% of passwords contained at least one symbol. The comparison between MySpace passwords and passwords which our participants had before and after the education is shown in Figure 1.

In the rest of the paper we use our categorization. With these, 4,3% of passwords were Word passwords, 51,1% were Mixture passwords and the rest 44,7% were Non-word passwords. The password structures before the education session in our study is shown in Table IV in columns “before”. In 2008, most of word-parts in the Mixture passwords were modified with uppercase letters remaining word part readable. In 2009, also digits were used together with uppercase letters in word-part modification. Otherwise password generation process was rather similar in both groups.

One week after education (classroom education and home study), the distribution of the password structure and designing process had changed. None of passwords were pure words. One student (2,9%) had used Word-password structure where all words were modified. For the modification all character sets were used. Mixture passwords were 26,5% of all passwords and in 77,8% of them characters from all character sets were used. From all passwords 70,6% were Non-word passwords and 91,7% of these consist of characters from all character sets. The distribution of the password structure after the education is shown in Table IV in columns “after”.

Most of students (81,3%) in the classroom study group (group 1) selected Non-word password after class education. Mixture password was only chosen by 1,3% of students. However, in the self-study group (group 2), the percent of Mixture passwords was 38,9%, while the rest of the passwords were Non-word passwords. Comparing

TABLE IV.
PASSWORD STRUCTURES AND GENERATION PROCESS BEFORE THE EDUCATION (BEFORE), ONE WEEK AFTER THE EDUCATION (AFTER), AND 6 MONTHS AFTER THE EDUCATION (6M). THE CLASSROOM STUDY GROUP IS G1 AND THE HOME STUDY GROUP G2. OTHER ABBREVIATIONS ARE FOLLOWING: LC: LOWERCASE LETTERS, UC: UPPERCASE LETTERS, D: DIGITS, S: SPECIAL CHARACTERS, ALL: ALL CHARACTER SETS, W: WORD, MOD:ALL: MODIFICATION WITH ALL CHARACTER SETS AND NWP: NON-WORD PART. PASSWORDS WITH "HUMAN SELECTED" STUDENTS CLAIMED TO CHOOSE CHARACTERS RANDOMLY. WITH "PASSPHRASE", A MNEMONIC SENTENCE IS USED.

Password Category		Before G1(23)	Before G2(24)	Sum (47)	After G1(16)	After G2(18)	Sum (34)	After 6m(8)
Word		0	2	2	1	0	1	0
	3 w		1	1				
	2 w		1	1				
	4 w, mod:all				1		1	
Mixture		12	12	24	2	7	9	4
	1 w, nwp:d	7	5	12				
	1 w, mod:uc, nwp:d							1
	1 w, mod:all, nwp:d				1		1	
	1 w, nwp:s	1		1				
	1 w, nwp:d,s	2	1	3				
	1 w, nwp:d,s					1	1	
	2 w, nwp:d	1	2	3				
	2 w, nwp:s				1		1	
	2 w, nwp:d,s	1	1	2		2	2	
	2 w, mod:d, nwp:d,s							1
	2 w, mod:uc, nwp:d,s							1
	2 w, nwp:uc,lc,d,s		2	2				
	2 w, mod:all, nwp:uc,d,s					1	1	
	2 w, mod:all, nwp:all					1	1	
	3 w, mod:d, nwp:s							1
	3 w, nwp:uc,lc,d		1	1				
	4 w, mod:all, nwp:s					1	1	
	4 w, mod:all, nwp:uc,lc,s					1	1	
Non-word		11	10	21	13	11	24	4
	Lc,d, Human selected	1		1				
	Uc,s, Alphabetical order				1		1	
	Uc,lc,d, Quickly typed		1	1				
	Uc,lc,d, Human selected	4	1	5				1
	Uc,lc,d, Passphrase	1	1	2		1	1	2
	Lc,d,s, Human selected		1	1				
	Uc,lc,d, Computer gen.	2	5	7				
	All, Human selected		1	1	5	6	11	1
	All, Quickly typed	2		2	1	1	2	
	All, Passphrase	1		1	5	2	7	
	All, Computer gen.				1	1	2	

groups, the group 2 had stretched the variety of the password structure more than the group 1. The large variety of password structures can be thought as a defence against targeted attacks. Before the education an adversary could have done an educated guess successfully based on, for example findings in [12]. After the education the password structures had changed significantly and guessing had become meaningless.

The Table IV shows also the structures of the passwords students had half a year after the education. Half of the students used Mixture passwords and half Non-word passwords. Comparing word parts of the Mixture passwords before and half a year after, it can be seen the education had a slight positive effect. The word parts in the Mixture passwords were indeed modified. However, the special characters were not used in modification as was done one week after the education. The same phenomenon was seen in Non-word passwords. Only one student had used all four character sets when generating the password. Others had only used alpha-numeric characters.

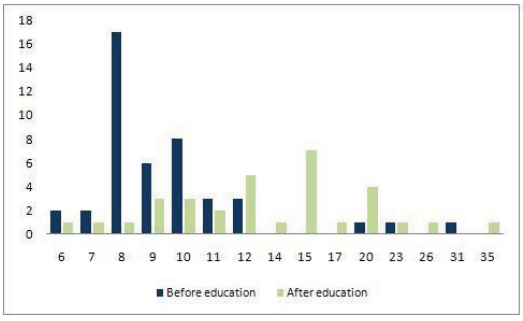


Figure 2. Length distributions

B. Password Length

In MySpace data [12] the most popular password lengths were 6 (15%), 7 (23%), 8 (25%), and 9 (17%) characters. In our case the results were rather similar. 68,7% of passwords had length between 8 and 10 charac-

ters. The most popular length was 8 characters (37,8%). The password length distribution before education is shown in Figure 2. Average password length was 9,0 characters for the group 1 (2008), 10,8 characters for the group 2 (2009), and 9,9 characters for both groups together.

After education session, the password length had become more evenly distributed and also the average password length had gotten higher, see Figure 2. This is a good result in two ways. First, the individual passwords have become stronger based on the longer passwords, and the second, the whole password group has become stronger. This means that the adversary knowing the password policy, does not benefit from it. For example, in our University College the password policy states that "minimum password length is 8 characters". Before the education, the adversary had 37,8% chance that student password was 8 characters long. After education, only 5,9% of students had 8 character long password. When looking groups separately, it can be noticed that gap between groups have gotten larger than it was before education. Before education, the group 2 had 1,8 characters longer passwords, and after the education the difference is 4,7 characters.

The average of the password length of the group 2 after half a year from the education had not dropped to the "before education" -level being now 12,1 characters. When comparing passwords with their category guidelines, the passwords were heading to the right directions. Non-word passwords were 8 or more characters long (guideline states "longer than 8" characters), and Mixture passwords were 10 or more characters long (guideline states "longer than 10" characters).

C. Password Strength

When measuring the strength of passwords with our tool, two thresholds were used. Passwords achieving more than 735 points were characterized as good passwords, while passwords more than 875 points were characterized as strong passwords.

Before education the passwords were in general weak. In group 1, 78% of the passwords were weak, the median score was 476 points, and the mean was 485 ± 144 points with 95% confidence interval. In group 2, 94% of the passwords was weak, the median was 456 points, and the mean was 585 ± 340 with 95% confidence interval. After the education passwords had gotten stronger. Among the group 1, 20% had a good password and 53% a strong password. The median score was 945 points and the mean was 952 ± 202 points with 95% confidence interval. In the group 2, the percents of the good passwords were 19% and strong passwords 75%. The median was 1234 points and the mean 1262 ± 242 points with 95% confidence interval. This strengthening is due to changes in the password structures, use of all character sets and longer passwords.

Group 2 has higher password quality points both before and after education than group 1. This can be explained with the length of the passwords. It is also observed that group 2 has always had longer passwords than the group 1.

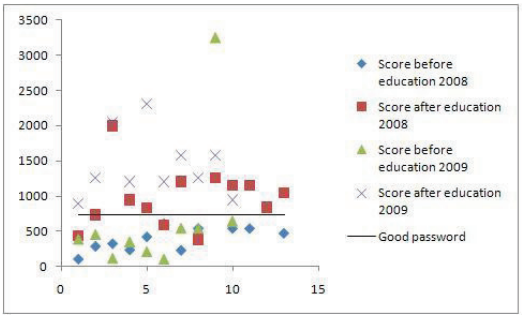


Figure 3. Scores before and after education with individual changes

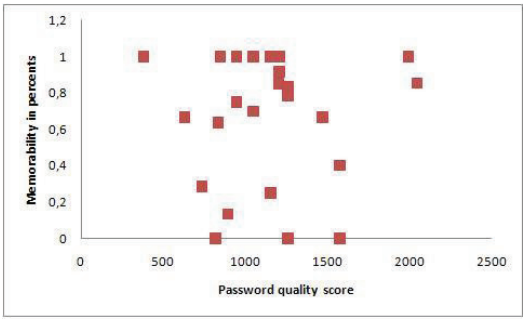


Figure 4. Memorability vs. Quality Score

Figure 3 shows how the password strength changed individually. The data shown in the figure consist of the students from both groups who delivered two first questionnaires successfully.

Long term educational effect was measured half a year after the education among the group 2 students. We noticed that the positive effect of the education, which was so obviously seen after one week from the education, had vanished. In most cases password were weak again. The median had dropped to 544 points and the mean to 590 ± 205 points with 95% confidence interval. The most important reason for strength decreasing was the lack of special characters.

D. Memorability

After a week from password designing date, students were asked to write down their passwords. Students were told not to use their passwords in any applications because the passwords were required to reveal. 48,2% of those who delivered third questionnaire, remembered password either correctly or with 1-2 easily corrected errors. With easily corrected errors we mean errors that are noticed when typed, for example that a password was one character too short. We did not use computers to collect passwords. All questionnaires were on paper and

students filled them in a lecture. Taking into consideration that writing a password on a paper is a different process that typing a password on the keyboard, we chose to take into account also 1-2 character errors while password still being a memorabable password.

Here we also find a difference between groups. The group 1 remembered their passwords better than the group 2. Percents being respectatilively 61,5% and 37,5%. This can be due to the fact that the passwords used by the group 1 were shorter than the passwords used by the group 2.

Figure 4 shows the relations between password memorability and strength. Memorabilty is computed as percents of right characters in a password string. It can be seen that the strong passwords are also memorable.

Zviran and Haga have studied password memorability in several occasions. While studying memorability of cognitive passwords, they used self-generated and system-generated passwords as control group. The recall rates for self-generated passwords after three months were 31% and system-generated 24% [15]. In [16] recall rate for self-selected passwords after three months were 35% and for random passwords 23% All passwords recalled in our study, except one, were self-selected. That one password was computer generated but the student had made the program by himself.

V. Discussion and Future Work

Because the password categorization was new to the students, it had to be thoroughly explained to the students in order to get the password quality measurements correct. However, despite careful explanation, some of the students still find it hard to answer correctly.

We also had to exclude some of the deliveries because the users had used other language than Norwegian. We did not have other languages included our measurement tool, and therefore the password quality scores were not accurate. However, similar instructions as given in password guidelines can be applied for other languages and the passwords made based on the guidelines are good or strong passwords.

The number of participants in our study became progressively less to the end. The number of acceptable answers was too small in the end to make the firm decision of which teaching method is the most efficient. It would be interesting to see how the password strength would change if the password guidelines shown in this paper were always visible to the users when changing a password. This would remind users of good password generation habits, and the effect of the continuous education could be measured.

VI. Conclusion

In this study, we taught two student groups how to design strong passwords in three different password categories: Word passwords, Mixture passwords and Non-word passwords. The education differed between groups. One group received a classical class room lecture session and the second studied the same material at home as a self study. In both cases, the passwords designed right after the education were much stronger. Not only password length got longer, but also all character sets were taken into password designing processes.

The most important finding after the password education was that password structure changed. The variation among structures had increased, and the distribution of password structure and also password length had become closer to the uniform distribution. The large variety of password structures and lengths is a good defence against targeted attacks.

However, one-time education did not help students to generate good passwords after half a year from the education. Within this timeframe the password guidelines of the high school had not changed and student had not received any remainders of strong passwords. We asked the home study group to answer to the fourth questionnaire. Only eight students answered. Despite the small number of delivering, a tendency could be derived. The students had problems to add special characters to their passwords. Passwords were also slightly shorter than right after the education. Both these reasons influenced on password strength by making it weak again.

In order to make firm predictions of which teaching style gives better results, the number of students in each group should have been higher. However, based on the results in this study, the self study group achieved larger password structure variation, longer passwords and higher quality points. This indicates that the self study is as sufficient as a traditional class teaching.

When considering memorability, the study supports the idea that self generated passwords are easier to remember than computer generated. In our study all remembered passwords, except one, were self generated. The study also indicates that passwords made based on different policies among different password categories are at least as memorable as other passwords.

Acknowledgements

The author is grateful to the students participating this study, to Nils Kalstad Svendsen for guiding the self study group and to Einar Snekkenes for valuable feedback. This work is supported by the Research Council of Norway, grant 158777, Authentication in a health service context.

References

[1] A. Brown, E. Bracken, S. Zoccoli, and K. Douglas. Generating and Remembering Passwords. *Applied Cognitive Psychology*, 18:641–651, 2004.

[2] E. Gehringer. Choosing Passwords: Security and Human Factors. In *Proceedings of ISTAS'02*, pages 39–373, 2002.

[3] K. Helkala and E. Snekkenes. A method for ranking authentication products. In *Proceedings of the International Symposium on Human Aspects of Information Security & Assurance (HAISA 2008), July 8-9, Plymount, UK*, pages 80–93, 2008.

[4] K. Helkala and E. Snekkenes. Password generation and search space reduction. *Journal of Computers*, 4, Issue 7:663–669, 2009.

[5] Kirsi Helkala. An educational tool for password quality measurements. In *Proceedings of Norwegian Information Security Conference*, pages 69–80. Tapir Akademisk Forlag, 2008.

[6] Ann-Marie Horcher and Gurvirender P. Tejay. Building A Better Password: The Role of Cognitive Load in Information Security Training. In *Proceedings of IEEE International Conference on Intelligence and Security Informatics, ISI*, pages 113–118, 2009.

[7] Information Technology Services, The Pennsylvania State University. Password Policy. `its.psu.edu/policies/password.html`, 2006.

[8] ISO. *NS-ISO/IEC 17799:2001*.

[9] Frank H. Katz. The Effect of a University Information Security Survey on Instruction Methods in Information Security. In *Proceedings on Information Security Curriculum Development Conference*, pages 43–48, 2005.

[10] C. Kuo, S. Romanosky, and L.F. Cranor. Human selection of mnemonic phrase-based passwords. In *Proceedings of the second symposium on Usable privacy and security*, volume 149 of *ACM International Conference Proceeding Series*, pages 67–78. ACM Press, 2006.

[11] M.A. Sasse, S. Brostoff, and D. Weirich. Transforming the "weakest link" - human/computer interaction approach to usable and effective security. *BT Technol*, 19(19):122–131, 2001.

[12] B. Schneier. Crypto-Gram Newletter: Real-World Passwords. `www.schneier.com/crypto-gram-0612.html`, December 2006.

[13] Robin Snyder. Ethical Hacking And Password Cracking: a Pattern For Individualized Security Exercises. In *Proceedings on Information Security Curriculum Development Conference*, pages 13–18, 2006.

[14] J. Yan, A. Blackwell, A. Anderson, and A. Grant. The Memorability and Security of Passwords - Some Empirical Results. Technical Report 500, Computer Laboratory, University of Cambridge, 2000.

[15] M. Zviran and W.J. Haga. User authentication by cognitive passwords: an empirical assessment. In *Proceedings of the 5th Jerusalem Conference on Information Technology*, pages 137–144, 1990.

[16] M. Zviran and W.J. Haga. A Comparison of Password Techniques for Multilevel Authentication Mechanisms. *Computer Journal*, 36(3):227–237, 1993.

Part V
Ranking Authentication Products

Chapter 15
A Method for Ranking Authentication Products

Kirsi Helkala and Einar Snekkenes

In *Proceedings of the International Symposium on Human Aspects of Information Security & Assurance (HAISA 2008)*, July 8-9, Plymount, UK, pp. 80-93, 2008

Chapter 16
Formalizing the Ranking of Authentication Products

Kirsi Helkala and Einar Snekkenes

In *Information Management & Computer Security - Special Issue*, Vol. 17, Issue 1, pp. 30-43, Emerald, 2009